Space-Frequency Localized Basis Function Networks for Nonlinear System Identification and Control

by

Mark Cannon

M.Eng., University of Oxford (1993)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1995

Author .....
Department of Mechanical Engineering
February 16, 1995

Certified by .....
Jean-Jacques E. Slotine
Associate Professor of Mechanical Engineering
and Information Sciences
Thesis Supervisor

Accepted by .....
Ain A. Sonin
Chairman, Departmental Committee on Graduate Students

Space-Frequency Localized Basis Function Networks for Nonlinear System Identification and Control

by

Mark Cannon

Submitted to the Department of Mechanical Engineering
on February 16, 1995, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

This thesis studies stable neural network approaches to estimation and control of nonlinear dynamic systems, and illustrates the discussion on a robotic motion-vision coordination application. Specifically, it extends earlier results on adaptive neuro-control using gaussian radial basis functions to the on-line generation of irregularly sampled networks, using tools from multiresolution analysis and wavelet theory, thus allowing much more compact and efficient system representations while preserving global stability.

An algorithm for stable, on-line adaptation of output weights simultaneously with node configuration in a class of non-parametric models with wavelet basis functions is presented. In addition, an asymptotic bound on the error in the network's reconstruction error is shown to be dependent on the residual approximation error associated with the steady state node configuration alone.

Using prior bounds on the temporal bandwidth of the controlled or identified system, an approach to reducing the redundancy of radial and ridge wavelet candidate basis function sets without compromising the efficiency of approximation is described. The problem of choosing which of the members of a non-orthonormal family to include in the approximation is solved for a class of wavelet using regularization principles. An experimental study of the use of network predictions of the path of an unknown object thrown through air to generate trajectories for an active-vision based robotic catching system is presented to illustrate the network's performance in practice.

Thesis Supervisor: Jean-Jacques E. Slotine
Title: Associate Professor of Mechanical Engineering
and Information Sciences

Acknowledgments

The author would like to thank Jean-Jacques Slotine for supplying a stream of original ideas, Ken Salisbury for coordinating experimental work, and Stephane Mallat for his helpful comments. Thanks also to Matt Williamson for his support, and Jesse Hong for his paper aeroplanes. Thanks to Zo and To for their patience and understanding.

This thesis was supported in part by NASA/JPL Contract No: 959774, "Vision and Touch Guided Grasping of Stationary and Moving Objects", and in part by a grant from Fujitsu. Development of the FEG camera actuators used in the experimental studies within this thesis was sponsored in part by a grant from the Sloan Foundation.

Contents

List of Figures

Chapter 1

Introduction

Most techniques for estimation and control of nonlinear systems with unknown dynamics rely on constructing nonparametric models from a finite history of input-output data. Given a sound theoretical framework, the practical success of such an approach depends both on the accuracy and efficiency of the chosen model's approximation, and on the extent to which these factors are governed by available prior information. This thesis studies the design and implementation of neural network architectures that provide efficient representations with little prior information through stable on-line adaptation of the model's structure concurrently with its output coefficients. Specifically, it extends earlier results on adaptive neurocontrol using gaussian radial basis functions [34, 33, 31, 35] to the on-line generation of irregularly sampled networks, using tools from multiresolution analysis and wavelet theory, thus allowing more compact and efficient system representations while preserving global stability.

We consider the identification and stable adaptive control of a class of time-invariant (or slowly varying) systems whose state $\mathbf{x} = [x\ \dot{x} \ldots x^{(n-1)}]^T$ evolves according to

$$x^{(n)}(t) + f[\mathbf{x}(t)] = u(t), \tag{1.1}$$

where the state $\mathbf{x}$ is measured, $u(t)$ is the control input (including the case when $u(t)$ is identically zero), and the function f is unknown. In both control and identification contexts, approximations to f are generated on-line using models that correspond to

single hidden layer neural networks:

$$\hat{f}(\mathbf{x}) = \sum_{j \in J} \hat{c}_j \phi_j(\mathbf{x}). \tag{1.2}$$

The restriction of particular types of nonlinearity to regions of the state space is likely to result in significant variations with position of the approximated function's local spatial bandwidth. In the dynamics of a robotic actuator for example, frictional effects are generally localized within regions of the state-space that correspond to low joint velocities. Owing to the rapid variation of low-speed frictional forces with position in state-space, f will be highly irregular in such regions.

These considerations suggest that very redundant representations are obtained using networks that have spatially uniform approximation capabilities. In order to reduce node densities within regions of the state space on which the data set varies smoothly, however, a detailed knowledge of the system is needed. Since such information is not generally available *a priori*, the network's structure must be adapted according to the *local* characteristics of the system during training in order to achieve an efficient representation.

In an estimation context, the centres and scalings of global activation basis functions such as sigmoids are often adjusted using gradient-descent methods with the objective of minimizing the sum of squared output errors at sample points alone. As a result of the global nature of their basis functions however, very large numbers of iterations are typically required before parameters of these networks converge. Moreover, the resulting network cannot be guaranteed to achieve the minimum reconstruction errors for a given number of nodes, the strongest result available being convergence to a local minimum of the error surface in parameter space, of which their may be many [4].

By contrast, the rapid convergence properties of expansions over spatially localized basis functions allow fast reconstruction and estimation of their linear coefficients. However, stable on-line adaptation of the centres and variances of the nodes of networks employing such basis functions is extremely difficult to achieve. The difficulty

arises as a result of the need to include a measure of the errors in these parameters explicitly in a Lyapunov stability analysis, whereas the effect on the network's reconstruction error of independently varying the characteristics of individual basis functions is highly nonlinear. To overcome this problem, [34, 33] study a network with centres constrained to lie on a regularly spaced grid, but whose spacing is adjusted on-line according to a time-varying estimate of the global bandwidth of f driven by a measure of control or estimation performance. Owing to the linear dependence of the network's output error bound on the error in the bandwidth estimate, laws for stable structural adaptation can be developed in this case. However, as discussed in these studies, the node configurations of such networks are not adapted to variations in the local frequency content of f, and a high degree of redundancy in the representations thus obtained is therefore unavoidable.

In order to benefit from both the fast training of linear output weights facilitated by the use of spatially localized basis functions and the increased efficiency afforded by structural adaptation, a means of estimating the local spatial frequency content of the approximated function that does not depend directly on reconstruction error is needed. Such an independent measure is available if basis functions that are localized simultaneously in space and frequency are employed. In this case, the magnitude of each coefficient in the expansion gives a direct measure of the content of the approximated function in the region of the space-frequency domain on which the energy of the associated basis function is concentrated. Structural adaptation is then possible through the selection of nodes on the basis of thresholding rules applied to coefficient magnitudes. This is the motivation for the space-frequency localized basis function networks described in this thesis.

Chapter 2 develops extensions of the adaptive neural network control and identification techniques described in [34, 33] to the case of space-frequency localized basis function models. Stable laws for concurrent adaptation of output weights and structure are derived and bounds on controller and estimator performance are given for general wavelet basis function families. The suitability of various classes of space-frequency localized basis function sets in control and identification applications are

investigated in chapter 3. Prior bounds on the temporal bandwidth of the system to be identified or controlled are used to develop a criterion for selecting radial and ridge wavelet basis functions for incorporation in the network, thus reducing the rate of increase in the network's size with the number of dimensions of the state vector. In chapter 4, the optimality properties of the algorithms for structural adaptation are explored in the context of wavelet and multivariate approximation theory. Chapter 5 describes an application of the network to the estimation of the trajectory of an unknown light bluff object moving freely through air. Experimental results are presented illustrating the performance of this technique. Chapter 6 presents brief concluding remarks.

Chapter 2

Network Architecture

This chapter studies the design of space-frequency localized basis function networks for control and identification applications. It motivates and describes an algorithm for selecting on-line the basis functions to be incorporated in the network, and shows that the coefficients in the resulting expansion can be concurrently estimated in a stable manner. It is assumed that a prior estimate is available of the supremum of the function f in the equation of state (1.1) on a compact subset of state space of interest.

The networks considered throughout this thesis consist of truncated expansions over families of wavelets. Expansions of this kind have recently found widespread use in signal processing and function approximation applications [8, 25, 21]. The basis functions of the network model are selected from a set $(\phi_j)_{j\in\mathcal{Z}^n}$ that is generated from a single function $\phi \in L^2(\mathcal{R}^n)$ via translations and dilations. The index $j \in \mathcal{Z}^n$ therefore consists of a scale or spatial frequency component p, and a translation component k.

A fundamental property of a set of candidate basis functions $(\phi_j)_{j\in\mathcal{Z}^n}$ for the networks developed in this thesis is that it constitute a *frame* for the class of functions to be approximated. Before detailing the network's design, the basic aspects of frame theory are first reviewed without proof. The condition that the family form a frame for $L^2(\mathcal{R}^n)$ is equivalent to the requirement that there exist finite and strictly positive

frame bounds m and M satisfying

$$m\|f\|_2^2 \leq \sum_{j\in\mathcal{Z}^n} |< f, \phi_j >|^2 \leq M\|f\|_2^2 \tag{2.1}$$

for all $f \in L^2(\mathcal{R}^n)$. It can be shown that the frame condition corresponds roughly to the requirement that the candidate basis function set completely covers the spatial and spatial frequency domains on which the approximated functions are defined, so that there exist no points $\mathbf{x}, \boldsymbol{\xi} \in \mathcal{R}^n$ for which $\phi_j(\mathbf{x}) = 0$ for all $j \in \mathcal{Z}^n$ or $\mathcal{F}\phi_j(\boldsymbol{\xi}) = 0$ for all $j \in \mathcal{Z}^n$ ($\mathcal{F}\phi$ denotes the Fourier transform of ϕ). This condition guarantees that the candidate basis function set spans the space of approximated functions, and that the network's reconstruction has low sensitivity to errors in estimated coefficients, as the following arguments show.

The inequality on the right side of (2.1) ensures that the frame operator F mapping functions $f \in L^2(\mathcal{R}^n)$ to sequences $(< f, \phi_j >)_{j\in\mathcal{Z}^n}$ is bounded. The frame adjoint operator F^*, defined for all $c = (c_j)_{j\in\mathcal{Z}^n} \in l^2(\mathcal{Z}^n)$ and $f \in L^2(\mathcal{R}^n)$ by

$$< F^*c, f >= \sum_{j\in\mathcal{Z}^n} c_j < \phi_j, f >$$

is then also bounded, since $\|F^*c\| \leq M\sum_{j\in\mathcal{Z}^n} c_j^2$. The definition of the frame adjoint operator also implies that $F^*c = \sum_{j\in\mathcal{Z}^n} c_j\phi_j$ (in the sense of L^2-convergence). Since $\|F^*\|$ is bounded, it follows that expansions of the form $\sum_{j\in\mathcal{Z}^n} c_j\phi_j$ can be computed in a numerically stable way, with small changes in the coefficients $(c_j)_{j\in\mathcal{Z}^n}$ corresponding to small changes in the reconstructed function. As a result, the frame bound M ensures that square-summable errors in a set of estimated coefficients cause a square-integrable error in the function reconstructed by an expansion over $(\phi_j)_{j\in\mathcal{Z}}^n$.

The left side of (2.1) ensures that the operator F^*F has a bounded inverse, so that a family of functions $(\tilde{\phi}_j)_{j\in\mathcal{Z}^n} \in L^2(\mathcal{R}^n)$ can be constructed using $\tilde{\phi}_j = (F^*F)^{-1}\phi_j$ for all j. This set of functions is itself a frame, known as the *dual* frame, with frame

bounds M^{-1} and m^{-1}. The formula

$$c_j = < f, \tilde{\phi}_j > \tag{2.2}$$

yields for all $f \in L^2(\mathcal{R}^n)$ a set of coefficients that exactly reconstruct f in an expansion over $(\phi_j)_{j \in \mathcal{Z}^n}$ [8, 21]. Prior bounds on the coefficients in a network expansion and bounds on the network's approximation capability can then be derived from bounds on the inner products (2.2).

The families with frame bounds $m = M = 1$ form an orthonormal basis for $L^2(\mathcal{R}^n)$, and the case $m = M \neq 1$ corresponds to the candidate basis function set constituting a so-called *tight* frame. The suitabilities of frames of orthonormal and non-orthonormal functions, and the construction of frames for classes of function defined on $\mathcal{R}^n$ are discussed in chapter 3. The component functions of tight and more general frames are typically not linearly independent. This means that for a given f, there may exist many different linear combinations of the ϕ_j that represent f exactly. The optimality properties of the algorithm for adapting the network's basis function set in both orthonormal and non-orthonormal cases are discussed in chapter 4.

2.1 Structural Adaptation

The preceding discussion shows that any function in $L^2(\mathcal{R}^n)$ can be approximated with arbitrary precision by an expansion over a set of functions $(\phi_j)_{j \in \mathcal{Z}^n}$ with finite frame bounds. In order to limit the number of nodes in the network expansion however, a finite number of basis functions must be selected from the frame that forms the candidate basis function set. This section describes how the effective truncation of the approximated function to a compact subset of state space allows the candidate basis function set to be restricted to a set of nodes with spatial centres lying on a bounded set. An algorithm for selecting recursively the network's nodes from the members of this truncated set on the basis of coefficient estimates is given, and bounds on coefficient magnitudes used to further restrict the candidate basis function set to

nodes localized at points within a compact set on the frequency domain.

In the context of control, the system (1.1) is to be forced to follow a desired trajectory $x_d(t)$, assumed to be bounded, and n times continuously differentiable, with bounded derivatives up to order n. The desired trajectory is therefore constrained to lie entirely within a bounded subset A_d of state-space. As described in section 2.2, a sliding controller is used to force the system's state back into A_d should it leave this set during training, and the function reconstructed by the network is consequently evaluated only on a compact set $A \supset A_d$. Alternatively, in the case of identification, the system's state is assumed to remain within a bounded set A_d at all times. As a result, the domain on which the network model is required to approximate the function $f(\mathbf{x})$ in the system's equation of state (1.1) is in this case restricted to the set $A = A_d$.

To allow approximation of functions with non-zero mean on A, a constant bias term is incorporated in the network model in addition to the wavelet basis functions. If the mean value of $f(\mathbf{x})$ on A is c, then without loss of generality, the function approximated by the component of the network that consists of an expansion over a set of wavelets may be assumed to be identical to $f-c$ on A but supported on a bounded set $A_T \supset A$. In order to avoid introducing artificially low rates of decay across scale in the coefficients of an exact reconstruction of the truncated function, the truncation of $f-c$ to A_T must preserve the smoothness properties of $f-c$ on A_T. This can be achieved by modulating $f(\mathbf{x}) - c$ by the function $m_T(\mathbf{x})$, where m_T is a member of the space $\mathcal{D}$ of infinitely differentiable functions that are identically zero outside of some compact set. Since $\mathcal{D}$ is dense in the space of continuous functions with compact support, the size of the region $A_T - A$ on which $f-c$ is truncated to zero is arbitrary, and m_T can be defined

$$\begin{aligned} m_T(\mathbf{x}) &= 1 && \text{if } \mathbf{x} \in A \\ 0 < m_T(\mathbf{x}) &< 1 && \text{if } \mathbf{x} \in A_T - A \\ m_T(\mathbf{x}) &= 0 && \text{if } \mathbf{x} \notin A_T. \end{aligned} \tag{2.3}$$

Since $f(\mathbf{x}) = m_T(\mathbf{x})(f(\mathbf{x}) - c)$ for all $\mathbf{x} \in A$, the function f_A approximated by

the space-frequency localized component of the network's model may be defined $f_A(\mathbf{x}) = m_T(\mathbf{x})(f(\mathbf{x}) - c)$.

It is shown in [8] that any compactly supported function in $L^2(\mathcal{R}^n)$ can be approximated with arbitrarily small L^2-error by an expansion over the functions $(\phi_j)_{j\in J}$ that have centres of spatial localization within a set $B(J)$ through choice of sufficiently large $B(J)$, provided ϕ is well-localized in space and frequency, and $(\phi_j)_{j\in\mathcal{Z}^n}$ has finite frame bounds. This result suggests that an accurate representation of f_A can be obtained with basis functions selected from the subset of the family $(\phi_j)_{j\in\mathcal{Z}^n}$ that consists of functions with centres lying within A_T.

Let the union J of disjoint index sets J_t^+, J_t^-, and J_t^0 contain the indices of all nodes whose centres lie at points included in A_T. The infinite set J_t^0 consists of the indices of nodes that are neglected in the expansion at time t, so that the function reconstructed by the network can be written

$$\hat{f}(\mathbf{x}) = \sum_{j\in J_t^+\cup J_t^-} \hat{c}_j(t)\phi_j(\mathbf{x}) + \hat{c}(t), \tag{2.4}$$

The network's structure is adapted by continually updating the set of nodes that are incorporated in the reconstruction $\hat{f}$ according to the following laws:

- Start with a single node in the network (in addition to the bias term $\hat{c}$), with support containing the initial position of the state vector $\mathbf{x}(0)$.
- Select a node with index j for removal from the network if $j \in J_t^-$ at time t.
- Introduce a node at time t by adding its associated index to the set J_t^+.

The membership of sets J_t^+, J_t^-, and J_t^0 is determined by simple thresholding rules applied to the magnitudes of nodes' coefficients, as will now be discussed.

A node is selected to be removed from the network if the magnitude of its coefficient falls below a threshold λ during training. Similar criteria are used in truncating wavelet expansions in the adaptive approximation schemes proposed in [10, 2, 11, 12]. In order to allow the use of initial conditions $\hat{c}_j = 0$ in the estimates of coefficients of nodes newly introduced into the network, the rule for membership of J_t^- includes

the condition that the derivative of a coefficient's magnitude is negative or zero. The set J_t^- is therefore defined by

$$J_t^- = \{j; |\hat{c}_j(t)| \leq \lambda \text{ and } \frac{d}{dt}\hat{c}_j^2(t) \leq 0\}.$$

The threshold λ can also be used to exclude functions from the candidate basis function set on the basis of prior bounds $|c_j|_{\max}$ on their estimated coefficients. Any basis function with an associated coefficient bound smaller than λ will remain in the set J_t^+ only if the magnitude of its estimated coefficient is too large. This suggests that only nodes with coefficient bounds exceeding λ should be admissible as basis functions in the network expansion, so that the set J_t^+ is defined by

$$J_t^+ = \{j; |\hat{c}_j(t)| > \lambda \text{ or } \frac{d}{dt}\hat{c}_j^2(t) > 0, \text{ and } |c_j|_{\max} > \lambda\}.$$

Bounds $|c_j|_{\max}$ on the magnitudes of coefficients in an expansion of f_A over the set $(\phi_j)_{j\in\mathcal{Z}^n}$ can be estimated using the formula (2.2). Assuming an estimate of $|f|_{\max} = \sup_{\mathbf{x}\in A_T} |f(\mathbf{x}) - c|$ is available, and that the elements of the dual frame are bounded, appendix 2.7 derives the following bounds

$$|c_j|_{\max} = a_0^{-n|p|/2} |f|_{\max} |\tilde{\phi}|_{\max} |A_T|. \tag{2.5}$$

In this expression, $|p|$ is the modulus of the scale parameter associated with the index j, $|A_T| = \int_{A_T} dx$ is the volume of A_T, and $|\tilde{\phi}|_{\max} = \sup_{\mathbf{x}\in A_T, k\in\mathcal{Z}^n} |\tilde{\phi}_{j(0,k)}(\mathbf{x})|$ where $j(p,k)$ is the index with dilation parameter p and translation parameter k. These bounds are valid for the coefficients in expansions over n-dimensional separable, ridge and radial wavelet families with scale parameter a_0.

The requirement that $|c_j|_{\max} > \lambda$ imposes the condition $|p| < p_\lambda$ on the scales of all nodes incorporated in the network expansion, where

$$p_\lambda = \frac{2}{\log a_0} \log\left(\frac{|f|_{\max} |\tilde{\phi}|_{\max} |A_T|}{\lambda}\right). \tag{2.6}$$

This condition further restricts the candidate basis function set to a finite number of functions that are localized at points lying within a bounded set on the spatial frequency domain. Since a reduction in the value of λ causes p_λ to increase, making basis functions with higher spatial frequency content available for inclusion in the network, the relationship between λ and p_λ agrees with the intuition that the effect of reducing λ is to increase the approximation capability of the network (see figure 2-1). Furthermore, $dp_\lambda/d\lambda$ is inversely proportional to λ, so that the sensitivity of p_λ to changes in λ increases as λ is reduced. This reflects the need for increasingly large numbers of nodes in order to achieve a given degree of improvement in approximation accuracy as the network's approximation errors become small. A method of determining the value of λ so as to balance the desired accuracy of approximation with limits on the numbers of network nodes that result from finite computational resources is discussed in section 2.5.

The criterion for selecting nodes to add to the network is based on a heuristic motivated by the degree of correlation between functions occupying neighbouring positions on the lattice of basis functions' space-frequency localization points. A basis function with index j is admissible for inclusion in the network's expansion if $j \in J_t^0$ and $|c_j|_{\max} > \lambda$. New nodes are introduced at points in the space-frequency lattice immediately adjacent to the locations of nodes whose coefficients exceed a threshold μ in absolute value, where μ is chosen so that $\mu/\lambda \ll 1$. The function of this parameter is to ensure that there is a small likelihood of a node with associated coefficient greater than λ not being introduced in the network. The magnitude of μ reflects the trade-off between the optimality of the steady-state node configuration and the computational cost of continually introducing large numbers of nodes into the network.

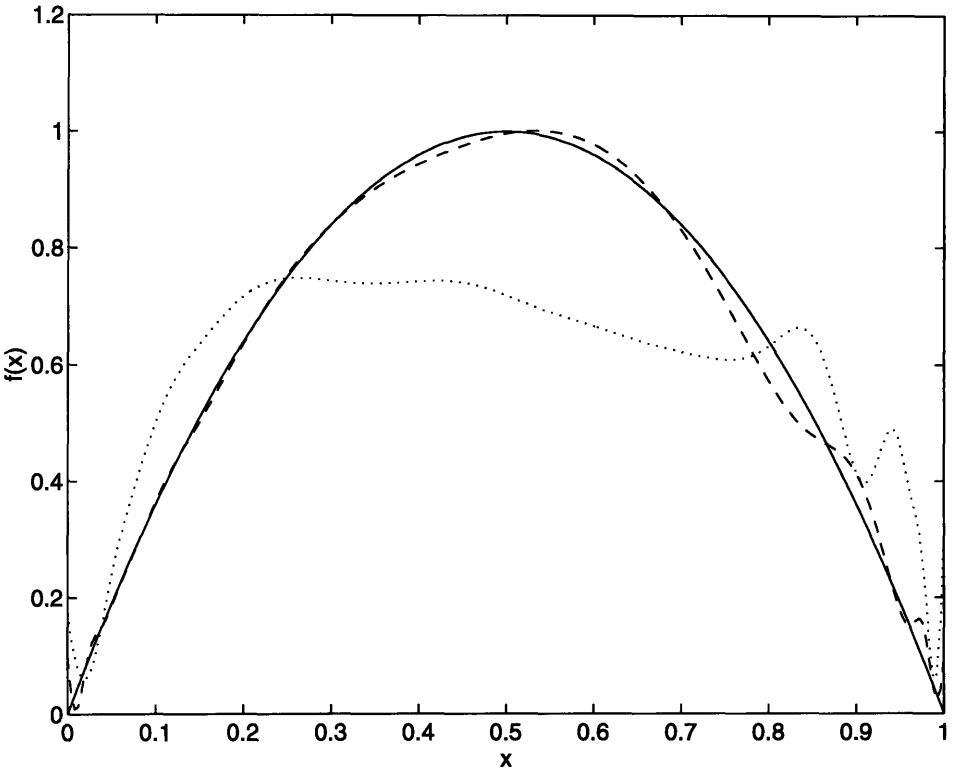

Figure 2-1: The identification wavelet network trained on the chaotic time series generated by the logistic map $x[t+1] = 4x[t](1-x[t])$: steady-state approximations obtained using $\lambda = 0.075$ (broken line), $\lambda = 0.1$ (dotted line). With $\lambda = 0.075$, 18 nodes were obtained in the steady-state, and with $\lambda = 0.1$, 7 steady-state nodes were obtained. "Mexican hat" wavelets were used ($\phi(x) = (1-x^2)\mathrm{e}^{-x^2/2}$), and the values of p_λ were 8.60 and 7.78 for $\lambda = 0.075$ and $\lambda = 0.1$ respectively; the corresponding maximum moduli the scales of basis functions incorporated in the expansion were 2 and 4 respectively.

2.2 Stable Adaptive Controller Design

An adaptive tracking controller incorporating a space-frequency localized basis function network can be designed as a natural extension of the frameworks developed in [34, 33, 31, 35] for the design of stable radial gaussian basis function neurocontrollers. In this section, a direct adaptive controller for the nth-order unity gain system (1.1) is described, and output weight adaptation laws that are robust to modeling uncertainties and disturbances introduced by structural adaptation are given.

Define a tracking error metric

$$s(t) = \left(\frac{d}{dt} + \omega\right)^{n-1} \tilde{x}(t), \tag{2.7}$$

with $\tilde{x} = x - x_d$ representing the tracking error, and ω a strictly positive constant (see e.g. [37]). Using this measure of tracking error, the control law can be expressed

$$u(t) = -k_D s(t) + [1 - m_d(t)]u_n(t) + m_d(t)u_{sl}(t). \tag{2.8}$$

The component u_n feeds forward the network's approximation $\hat{f}$, and $k_D > 0$ is a p.d. gain. Since the network's approximation cannot be expected to be accurate for $\mathbf{x} \notin A_d$, a sliding component u_{sl} is included in the control law in order to force the system's state back into the set A_d in the eventuality of it leaving this nominal operating region during training [34]. The function m_d provides a smooth interpolation between adaptive and sliding modes of controller operation in the region between A_d and the bounded set $A \supset A_d$:

$$\begin{aligned} m_d(t) &= 0 && \text{if } \mathbf{x}(t) \in A_d \\ 0 < m_d(t) &< 1 && \text{if } \mathbf{x}(t) \in A - A_d \\ m_d(t) &= 1 && \text{if } \mathbf{x}(t) \notin A. \end{aligned}$$

Since the control signal contains no adaptive component if the state leaves the set A, the region of state-space on which the network model is used to approximate the system's dynamics is restricted to A.

The adaptive component of the control law can be expressed $u_n(t) = \hat{f}(\mathbf{x}) + \dot{x}_r(t)$, where the reference signal x_r is a collection of terms on the right side of equation (2.7) such that $s(t) = x^{(n-1)}(t) - x_r(t)$. Since $x_r(t)$ depends only on the desired trajectory and $x^{(k)}(t)$ for $k \leq n-2$, $\dot{x}_r$ can be fully computed from the measured state. The closed-loop dynamics of the error metric are therefore specified by

$$\dot{s}(t) = -k_D s(t) + (1 - m_d)[\hat{f}(\mathbf{x}) - f(\mathbf{x})] + m_d[u_{sl}(t) - f(\mathbf{x}) - \dot{x}_r(t)]. \tag{2.9}$$

Using the control law (2.8), it can be shown that adaptation laws for the coefficients of basis functions ϕ_j of the form $\dot{\hat{c}}_j = -k_a s(t)\phi_j(\mathbf{x})$ would ensure stability of the closed-loop system and provide asymptotically perfect tracking performance in the ideal case of a finite basis function set $(\phi_j)_{j \in J}$ giving an exact linear parameterization of f_A [37].

Owing to the generality of the function f in the control problem currently considered, the network (2.4) cannot be assumed to reconstruct f exactly, and a finite residual approximation error ϵ is in general unavoidable. Let $\tilde{c}_j = \hat{c}_j - c_j$ be the error in the estimate of $\hat{c}_j$ relative to c_j, where the set of coefficients $(c_j)_{j \in \mathcal{Z}^n}$ reconstructs the approximated function f_A in an expansion over the family $(\phi_j)_{j \in \mathcal{Z}^n}$ (the existence of one such set $(c_j)_{j \in \mathcal{Z}^n}$ is guaranteed by the family's finite frame bounds). Then ϵ is defined as the pointwise error in the function reconstructed by the network in the limit of convergence of errors in the estimates of coefficients indexed by J_t^+ and the bias term,

$$\begin{aligned} \epsilon(\mathbf{x}, t) &= \lim_{\substack{\tilde{c}_j \to 0, \tilde{c} \to 0 \\ j \in J_t^+}} (\hat{f}(\mathbf{x}) - f(\mathbf{x})) \\ &= \sum_{j \in J_t^-} \tilde{c}_j \phi_j(\mathbf{x}) - \sum_{j \in J_t^0} c_j \phi_j(\mathbf{x}) + d(\mathbf{x}), \end{aligned} \tag{2.10}$$

for all $\mathbf{x} \in A$. The term $d(\mathbf{x})$ quantifies the error in $\hat{f}(\mathbf{x})$ on A due to truncation (in space) of the candidate basis function set to nodes indexed by J.

In appendix 2.8, it is shown that $|\epsilon(\mathbf{x}, t)|$ is uniformly bounded on A as a result of the bounds on f, the generating wavelet ϕ, and the elements of the dual frame

$(\tilde{\phi}_j)_{j \in \mathcal{Z}^n}$. However, the finite disturbance of the closed loop system due to ϵ can cause instability in the certainty-equivalent output weight adaptation law given above, an overtraining-type phenomenon. The training algorithm can be made robust to this disturbance by halting adaptation should the network reach an estimate of the limit of its approximation capability, determined from an estimated upper bound on ϵ [34, 33].

The bound on ϵ developed in appendix 2.8 is extremely conservative however, and would result in poor tracking performance if used to provide robustness to the residual approximation error in the laws for output weight adaptation. Although more precise bounds on ϵ would require a degree of prior knowledge of f not usually available, adaptation laws robust to this disturbance can be developed using the upper bounds $|c_j|_{\max}$ on the magnitudes of the estimated coefficients given in equation (2.5). Output weight adaptation can then be made robust to the disturbance ϵ by introducing a decay mechanism into the coefficients that exceed the bounds on their magnitudes [35].

Output weight adaptation laws must also be made robust to disturbances introduced by structural adaptation. During removal from the network, the magnitude of a node's associated coefficient error may actually increase over time. For this reason, the coefficients indexed by the set J_t^- must decay gradually, with reduced rates of decay as the number of nodes being removed concurrently from the network increases.

The full stable output weight adaptation laws are as follows

$$\begin{aligned} \dot{\hat{c}}_j(t) &= \begin{cases} -k_a(1-m_d)\text{sign}(\hat{c}_j(t))/N(J_t^-) & \text{if } j \in J_t^- \\ -k_a(1-m_d)s(t)\phi_j(\mathbf{x}) - \sigma_j(t)\hat{c}_j(t) & \text{if } j \in J_t^+ \end{cases} \\ \dot{\hat{c}}(t) &= -k_a(1-m_d)s(t) - \sigma(t)\hat{c}(t), \end{aligned} \tag{2.11}$$

where $k_a > 0$ is an adaptation gain, and $N(J_t^-)$ is the number of nodes indexed by the set J_t^-. Since the number of candidate basis functions is finite, the value of $N(J_t^-)$ is necessarily bounded, and the adaptation law for coefficients indexed by J_t^- therefore constrains these coefficients to decay to zero in finite time. The functions $\sigma_j(t)$ and

$\sigma(t)$ provide a smooth modulation of the robust decay mechanism, and are defined

$$\sigma_j(t) = \begin{cases} 0 & \text{if } |\hat{c}_j(t)| < |c_j|_{\max} \\ \sigma_0 \left[|\hat{c}_j(t)|/|c_j|_{\max} - 1 \right] & \text{if } |c_j|_{\max} \leq |\hat{c}_j(t)| \leq 2|c_j|_{\max} \\ \sigma_0 & \text{if } |\hat{c}_j(t)| > 2|c_j|_{\max}. \end{cases} \tag{2.12}$$

The global stability of the adaptive controller can be shown using the Lyapunov-like function V_c:

$$V_c(t) = \frac{1}{2}s^2(t) + \frac{1}{2k_a}\left[\sum_{j\in J} \tilde{c}_j^2(t) + \tilde{c}^2(t)\right], \tag{2.13}$$

where $\hat{c}_j = 0$ if $j \in J_t^0$. The errors in coefficients indexed by the set $J_t^0 \subset J$ are included in this definition to ensure that $V_c(t)$ is continuous and $\dot{V}_c(t)$ is piecewise continuous. The existence of a set of square-summable coefficients $(c_j)_{j\in\mathcal{Z}^n}$ guarantees that $V_c(t)$ is finite if $s(t)$, $(\hat{c}_j(t))_{j\in J_t^+\cup J_t^-}$ and $c(t)$ are bounded. In section 2.4, bounds on $\dot{V}_c(t)$ are used to show that $V_c(t)$ is uniformly bounded, which implies that the tracking error measure $s(t)$ and the network's coefficients $(\hat{c}_j(t))_{j\in J_t^+\cup J_t^-}$ and $c(t)$ are uniformly bounded. Furthermore, the definition 2.7 of $s(t)$ as the input of a stable linear system whose output is $\mathbf{x}(t)$ allows the conclusion that the system's state $\mathbf{x}(t)$ is also uniformly bounded. Asymptotic bounds on the convergence of the tracking error metric are developed in section 2.5.

2.3 Identification Network Design

The design of stable space-frequency localized basis function networks for dynamic system identification applications is considered in this section. The system to be identified is assumed to be stable, n-th order, and of the class specified by (1.1) with state $\mathbf{x}(t)$ measurable and control input $u(t)$ identically zero. The network's reconstruction is given by equation (2.4), and its basis function set is adapted according to the laws described in section 2.1.

In order to obtain a measure of the error in the approximation $\hat{f}(\mathbf{x})$ given measurements of the state variables $x, \dot{x}, \ldots, x^{(n-1)}$ alone, the network is used to generate

a signal $\hat{x}_{n-1}(t)$ that corresponds to an estimate of $x^{(n-1)}(t)$. The estimate is propagated according to

$$\dot{\hat{x}}_{n-1}(t) = -k_D\, e(t) + \hat{f}(\mathbf{x}),$$

where $k_D > 0$ is a constant, and $e(t)$ is defined by $e(t) = \hat{x}_{n-1}(t) - x^{(n-1)}(t)$. Since the signal $e(t)$ is then related to the approximation error $\hat{f}(\mathbf{x}) - f(\mathbf{x})$ by an SPR linear system:

$$\dot{e}(t) = -k_D\, e(t) + \hat{f}(\mathbf{x}) - f(\mathbf{x}), \tag{2.14}$$

this signal can be used in coefficient adaptation laws as a measure of the error in the network's approximation [37].

The following adaptation laws allow stable training of output weights concurrently with adaptation of the network's structure

$$\begin{aligned} \dot{\hat{c}}_j(t) &= \begin{cases} -k_a \text{sign}(\hat{c}_j(t))/N(J_t^-) & \text{if } j \in J_t^- \\ -k_a e(t)\phi_j(\mathbf{x}) - \sigma_j(t)\hat{c}_j(t) & \text{if } j \in J_t^+ \end{cases}, \\ \dot{\hat{c}}(t) &= -k_a e(t) - \sigma(t)\hat{c}(t). \end{aligned} \tag{2.15}$$

The gain k_a controls the sensitivity of adaptation to changes in the error measure e, and $N(J_t^-)$ denotes the number of nodes in the set J_t^-. As in the case of the adaptive controller, the uniform boundedness of $N(J_t^-)$ guarantees that the coefficients indexed by J_t^- decay to zero in finite time. The robustness mechanisms employed are analogous to those developed for the adaptive control network described in the preceding section. Specifically, $\sigma_j(t)$ and $\sigma(t)$ are defined by (2.12), and provide robustness to the disturbance introduced by the network's finite residual approximation error ϵ given in equation (2.10).

The identifier's global stability of can be shown using the Lyapunov-like function V_e:

$$V_e(t) = \frac{1}{2}e^2(t) + \frac{1}{2k_a}\left[\sum_{j\in J} \tilde{c}_j^2(t) + \tilde{c}^2(t)\right]. \tag{2.16}$$

As in the Lyapunov-like function $V_c(t)$ constructed in section 2.2, errors in the coefficients indexed by the set $J_t^0 \subset J$ are included in $V_e(t)$ so that $\dot{V}_e(t)$ is piecewise

continuous, and the existence of a square-summable sequence $(c_j)_{j\in\mathcal{Z}^n}$ ensures that $V_e(t)$ is finite if $e(t)$, $(\hat{c}_j(t))_{j\in J_t^+\cup J_t^-}$ and $c(t)$ are bounded. It is shown in section 2.4 that $V_e(t)$ is uniformly bounded, which implies that $e(t)$, $(\hat{c}_j(t))_{j\in J_t^+\cup J_t^-}$, and $c(t)$ are uniformly bounded. Since the state variable $x^{(n-1)}(t)$ is bounded by assumption, the estimate $\hat{x}_{n-1}(t)$ is also uniformly bounded. In section 2.5, asymptotic bounds on the convergence of the error measure $e(t)$ are developed.

2.4 Stability Analysis

This section investigates the stability of the network-based adaptive controller described in section 2.2 and identifier of section 2.3. The stability analysis is approached by first deriving upper bounds on the derivatives of the Lyapunov-like functions V_c and V_e, and then showing that these bounds are necessarily negative if the error measures $s(t)$ and $e(t)$ or any of the coefficient estimates $\hat{c}(t)$ and $(\hat{c}_j(t))_{j\in J_t^+}$ become sufficiently large. The negative definiteness of $\dot{V}_c$ and $\dot{V}_e$ outside of compact sets on the parameter error spaces $(s, \tilde{c}, (\tilde{c}_j)_{j\in J_t^+})$ and $(e, \tilde{c}, (\tilde{c}_j)_{j\in J_t^+})$ is then used to show that V_c and V_e are uniformly bounded.

The following expression for $\dot{V}_e$ can be obtained from the error measure dynamics (2.14) of the identification network,

$$\dot{V}_e(t) = -k_D e^2(t) + e(t)(\hat{f}(\mathbf{x}) - f(\mathbf{x})) + \frac{1}{k_a}\left[\sum_{j\in J_t^+\cup J_t^-} \tilde{c}_j(t)\dot{\hat{c}}_j(t) + \tilde{c}(t)\dot{\hat{c}}(t)\right].$$

The expression (2.9) for the derivative of the tracking error metric and the control law (2.8), yield an upper bound on $\dot{V}_c(t)$

$$\dot{V}_c(t) \leq -k_D s^2(t) + (1 - m_d(t))s(t)(\hat{f}(\mathbf{x}) - f(\mathbf{x})) + \frac{1}{k_a}\left[\sum_{j\in J_t^+\cup J_t^-} \tilde{c}_j(t)\dot{\hat{c}}_j(t) + \tilde{c}(t)\dot{\hat{c}}(t)\right].$$

In terms of the approximation error ϵ defined at equation (2.10), the reconstruction

errors in the models used in both control and identification networks can be written

$$\hat{f}(\mathbf{x}) - f(\mathbf{x}) = \sum_{j \in J_t^+} \tilde{c}_j(t)\phi_j(\mathbf{x}) + \tilde{c}(t) + \epsilon(\mathbf{x}, t).$$

Using the adaptation laws (2.15) for the identification network's coefficients, $\dot{V}_e$ can therefore be bounded by

$$\begin{aligned} \dot{V}_e(t) \;\leq\; & -\frac{k_D}{2}e^2(t) + \frac{1}{2k_D}\epsilon^2(\mathbf{x}, t) - \frac{1}{k_a}\sum_{j \in J_t^+} \sigma_j(t)\tilde{c}_j(t)\hat{c}_j(t) \\ & -\frac{1}{k_a}\sigma(t)\tilde{c}(t)\hat{c}(t) + \frac{1}{N(J_t^-)}\sum_{j \in J_t^-} |\tilde{c}_j(t)|. \end{aligned} \tag{2.17}$$

The adaptation laws (2.11) for the parameters of the adaptive controller similarly allow $\dot{V}_c$ to be bounded by

$$\begin{aligned} \dot{V}_c(t) \;\leq\; & -\frac{k_D}{2}s^2(t) + \frac{(1-m_d)^2}{2k_D}\epsilon^2(\mathbf{x}, t) - \frac{1}{k_a}\sum_{j \in J_t^+} \sigma_j(t)\tilde{c}_j(t)\hat{c}_j(t) \\ & -\frac{1}{k_a}\sigma(t)\tilde{c}(t)\hat{c}(t) + \frac{1}{N(J_t^-)}\sum_{j \in J_t^-} |\tilde{c}_j(t)|. \end{aligned} \tag{2.18}$$

Since the function $m_d(\mathbf{x})$ satisfies $m_d(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathcal{R}^n$, (2.18) provides a bound on $\dot{V}_c$ identical to the bound (2.17) on $\dot{V}_e$, with $e(t)$ replaced by $s(t)$. This bound is therefore used to show the stability of both control and identification schemes, with V denoting either V_c or V_e, and $s(t)$ interchangeable with $e(t)$.

The errors in coefficients indexed by the set J_t^- satisfy $|\tilde{c}_j| \leq |c_j| + \lambda$ for all $j \in J_t^-$, and the square-summability of the coefficients $(c_j)_{j \in \mathcal{Z}^n}$ ensures that $\sup_{j \in J_t^-} |c_j|$ exists. As a result, $\dot{V}(t)$ can be further bounded by

$$\begin{aligned} \dot{V}(t) \;\leq\; & -\frac{k_D}{2}s^2(t) + \frac{1}{2k_D}\epsilon^2(\mathbf{x}, t) - \frac{1}{k_a}\sum_{j \in J_t^+} \sigma_j(t)\hat{c}_j(t)\tilde{c}_j(t) \\ & -\frac{1}{k_a}\sigma(t)\hat{c}(t)\tilde{c}(t) + \lambda + \sup_{j \in J_t^-} |c_j|. \end{aligned} \tag{2.19}$$

Each of the terms in (2.19) containing the modulation function $\sigma_j(t)$ satisfies a

bound

$$\begin{aligned} -\sigma_j(t)\hat{c}_j(t)\tilde{c}_j(t) &= -\sigma_j(t)(\hat{c}_j^2(t) - c_j\hat{c}_j(t)) \\ &\leq -\sigma_j(t)|\hat{c}_j(t)|(|\hat{c}_j(t)| - |c_j|). \end{aligned}$$

Furthermore, it can be seen from the definition (2.12) that $\sigma_j(t)$ is non-zero only if $|\hat{c}_j| > |c_j|$, which implies that $-\sigma_j(t)\hat{c}_j\tilde{c}_j(t) \leq 0$. The inequality (2.19) therefore reduces to

$$\dot{V}(t) \leq -\frac{k_D}{2}s^2(t) + d_0(\mathbf{x}, t), \tag{2.20}$$

where $d_0(\mathbf{x}, t) = \epsilon^2(\mathbf{x}, t)/2k_D + \lambda + \sup_{j\in J_t^-} |c_j|$. It follows that a sufficient condition for $\dot{V}(t) < 0$ is $s^2(t) > 2d_0(\mathbf{x}, t)/k_D$.

An alternative bound on each of the terms in (2.19) that are modulated by $\sigma_j(t)$ is given by

$$\begin{aligned} -\sigma_j(t)\hat{c}_j(t)\tilde{c}_j(t) &= -\frac{\sigma_j(t)}{2}\tilde{c}_j^2(t) - \frac{\sigma_j(t)}{2}(\hat{c}_j^2(t) - c_j^2) \\ &\leq -\frac{\sigma_j(t)}{2}\tilde{c}_j^2(t), \end{aligned}$$

(where $\sigma_j(t) \neq 0$ only if $|\hat{c}_j(t)| > |c_j|$ has been used). For all $j \in J_t^+$, $\dot{V}(t)$ is therefore subject to the bound

$$\dot{V}(t) \leq -\frac{1}{2k_a}\sigma_j(t)\tilde{c}_j^2(t) + d_0(\mathbf{x}, t), \tag{2.21}$$

with $d_0(\mathbf{x}, t) = \epsilon^2(\mathbf{x}, t)/2k_D + \lambda + \sup_{j\in J_t^-} |c_j|$. The definition (2.12) of $\sigma_j(t)$ shows that $\sigma_j(t)$ is necessarily equal to σ_0 if $|\tilde{c}_j(t)| > |c_j|_{\max}$. From the inequality (2.21), a sufficient condition for $\dot{V}(t) < 0$ is therefore given for all $j \in J_t^+$ by
$\tilde{c}_j^2(t) > \max\{k_a d_0(\mathbf{x}, t)/2\sigma_0,\ 9|c_j|_{\max}^2\}$. A similar argument applied to the error in the bias term shows that $\dot{V}(t)$ is necessarily negative if $\tilde{c}^2(t) > \max\{k_a d_0(\mathbf{x}, t)/2\sigma_0,\ 9|c|_{\max}^2\}$.

The disturbance $d_0(\mathbf{x}, t)$ is uniformly bounded since the residual error $\epsilon(\mathbf{x}, t)$ is uniformly bounded and the coefficients $(c_j)_{j\in\mathcal{Z}^n}$ are square-summable. The conditions on the sign of $\dot{V}$ derived above therefore establish the existence of an $r_0 > 0$ so

that $\dot{V} < 0$ if $r > r_0$, where r is a radial measure on the parameter error space $(s, \tilde{c}, (\tilde{c}_j)_{j \in J_t^+})$, defined by

$$r^2(t) = s^2(t) + \sum_{j \in J_t^+} \tilde{c}_j^2(t) + \tilde{c}^2(t).$$

Since in addition, V is positive definite and decrescent in r, it follows that $V(t)$ is uniformly bounded.

Alternative coefficient adaption laws Several variations on the coefficient adaptation laws (2.11) and (2.15) are permissible within the framework of the stability analysis given above. A simple alternative to the robust decay mechanism introduced in coefficient estimates that exceed the bounds $|c_j|_{\max}$ is to saturate coefficient estimates at these bounds. The coefficient adaptation laws for the adaptive controller described in section (2.2) are in this case given by

$$\dot{\hat{c}}_j(t) = \begin{cases} -k_a(1-m_d)\text{sign}(\hat{c}_j(t))/N(J_t^-) & \text{if } j \in J_t^- \\ -k_a(1-m_d)s(t)\phi_j(\mathbf{x}) & \text{if } |\hat{c}_j(t)| < |c_j|_{\max} \text{ and } j \in J_t^+ \\ 0 & \text{if } |\hat{c}_j(t)| = |c_j|_{\max} \text{ and } j \in J_t^+, \\ & \quad \text{and } \text{sign}(s(t)\phi_j(\mathbf{x})) = -\text{sign}(\hat{c}_j(t)) \end{cases}$$

$$\dot{\hat{c}}(t) = \begin{cases} -k_a(1-m_d)s(t) & \text{if } |\hat{c}(t)| < |c|_{\max} \\ 0 & \text{if } |\hat{c}(t)| = |c|_{\max} \text{ and } \text{sign}(s(t)) = -\text{sign}(\hat{c}(t)), \end{cases}$$

and adaptation laws for the identification network's coefficients are defined analogously. The stability of the network-based controller and identifier can be shown using the the Lyapunov-like functions V_c and V_e of equations (2.13) and (2.16). Upper bounds on $\dot{V}_c$ and $\dot{V}_e$ are given by (2.20), so that $\dot{V}_c$ and $\dot{V}_e$ are again negative definite for sufficiently large error measures $|s(t)|$ and $|e(t)|$. Since the estimates $(\hat{c}_j)_{j \in J_t^- \cup J_t^+}$ are uniformly bounded and the coefficients $(c_j)_{j \in \mathcal{Z}^n}$ are square summable, it follows that $V_c(t)$ and $V_e(t)$ are uniformly bounded. Bounds on the convergence of errors developed in section 2.5 also hold in this case. This modification therefore simplifies the update equations for coefficient estimates without altering either the network's

stability or convergence properties.

Another modification allows stable adaptation laws that depend only on local system properties. The adaptation laws (2.11) and (2.15) require a knowledge of the total number of nodes indexed by the set J_t^-. If basis functions have compact support however, the output weights of nodes selected for removal from the network can be adapted on the basis of a knowledge of the set $J_{\mathbf{x},t}^- \subset J_t^-$ containing the indices of basis functions whose support includes $\mathbf{x}(t)$. Denoting the number of nodes indexed by $J_{\mathbf{x},t}^-$ as $N(J_{\mathbf{x},t}^-)$, adaptation laws for the set indexed by J_t^- can in this case be written

$$\dot{\hat{c}}_j(t) = -k_a(1-m_d)\frac{|\phi_j(\mathbf{x})|}{N(J_{\mathbf{x},t}^-)\sup_{\mathbf{x}\in A_T}|\phi_j(\mathbf{x})|}\text{sign}\hat{c}_j(t).$$

The generating wavelet is assumed bounded and the scales p of basis functions included in the network expansion are restricted by $|p| < p_\lambda$, so that $\sup_{x\in A_T}|\phi_j(\mathbf{x})|$ can be assumed to exist. The global stability of control and identification schemes incorporating this modified adaptation law can be shown using the argument given above, since the derivatives of the functions V_c and V_e obtained in this case again satisfy the bound (2.19). The local nature of the adaptation laws obtained with this modification would facilitate the design of parallel hardware implementations of the network.

2.5 Convergence Analysis

This section develops bounds on the the convergence of tracking errors for the adaptive controller described in section 2.2. These bounds have the same functional form as the available bounds on the convergence of the identification network's error measure. The tracking error metric s is therefore interchangeable with the identifier's error measure e in the following analysis.

The integral of equation (2.20) over an arbitrary interval T provides a bound on

the mean square tracking error,

$$\frac{1}{T}\int_t^{t+T} s^2(\tau)\,d\tau \leq \frac{1}{k_D^2}\sup_{\tau\in[t,t+T]}\epsilon^2(\mathbf{x}(\tau),\tau) - \frac{2}{k_D T}(V(t+T)-V(t)) + \frac{2}{k_D T}(\lambda + \sup_{j\in J}|c_j|).$$

Since $V(t)$ is bounded, and the coefficients $(c_j)_{j\in\mathcal{Z}^n}$ are necessarily bounded, the following asymptotic bound on the tracking error measure is obtained:

$$\limsup_{T\to\infty}\frac{1}{T}\int_t^{t+T} s^2(\tau)\,d\tau \leq \frac{1}{k_D^2}\sup_{\tau\in[t,t+T]}\epsilon^2(\mathbf{x}(\tau),\tau). \tag{2.22}$$

At any given time t_0, the tracking error metric can therefore be guaranteed to converge in a mean square sense to within a bound proportional to the supremum of the residual approximation error over times $t \geq t_0$. In contrast, the performance bounds for the radial gaussian neurocontrollers developed in [33] are obtained in a slightly stronger, pointwise sense. The bounds on the energy contained in the signals s and e that are available in this case are due to the use of bounds on the magnitudes of coefficient estimates, rather than bounds on the pointwise residual error, in the mechanisms that provide robustness to the disturbance caused by limits on the network's approximation capability.

Although finite, the pointwise bound on the network's residual error that is derived in section 2.8 is extremely conservative. This bound is based on estimated bounds on the magnitudes of coefficients indexed by the sets J_t^+ and J_t^-, and the estimate (2.6) of the maximum modulus of the scale of basis functions incorporated in the network expansion. In most cases of practical interest, the function $f(\mathbf{x})$ is uniformly continuous, or discontinuous at a finite number of points on A_T, so that the coefficient bounds (2.5) derived with no assumptions on the regularity of f, and estimates of the required bandwidth of the approximation based on p_λ, are extremely conservative. However, the value of ϵ is independent of the bounds $|c_j|_{\max}$, and the asymptotic bounds on the network's error measure (2.22) are therefore governed by the value of λ and the actual rates of decay of the coefficients $(c_j)_{j\in\mathcal{Z}^n}$.

A more meaningful measure of the residual approximation error is given by the L^2-norm of ϵ. The definition of $\epsilon(\mathbf{x},t)$ on A at equation (2.10) can be extended to

the whole of $\mathcal{R}^n$ as the residual error in the network's approximation to the function $f_A(\mathbf{x}) + c$:

$$\begin{aligned}\epsilon(\mathbf{x},t) &= \lim_{\substack{\tilde{c}_j \to 0, \tilde{c} \to 0 \\ j \in J_t^+}} \left[\hat{f}(\mathbf{x}) - (f_A(\mathbf{x}) + c)\right] \\ &= \sum_{j \in J_t^-} \tilde{c}_j \phi_j(\mathbf{x}) - \sum_{j \in J_t^0} c_j \phi_j(\mathbf{x}) - \sum_{j \notin J} c_j \phi_j(\mathbf{x}),\end{aligned}$$

for all $\mathbf{x} \in \mathcal{R}^n$. Since the coefficient sets $(\tilde{c}_j)_{j \in J_t^-}$, $(c_j)_{j \in J_t^0}$ and $(c_j)_{j \notin J}$ are square-summable, the boundedness of the adjoint frame operator can be used to bound the value of $\|\epsilon\|$,

$$\|\epsilon(\mathbf{x},t)\|^2 \leq M \left[\sum_{j \in J_t^-} \tilde{c}_j^2 + \sum_{j \in J_t^0} c_j^2 + \sum_{j \notin J} c_j^2\right],$$

where M is the upper frame bound, defined at (2.1). The second term in this expression gives a measure of the residual error due to truncating the network's expansion to basis functions centred on a compact set on the frequency domain, and is governed by the value of λ. The third term represents errors that result from the truncation of the candidate basis function set to basis functions with centres lying on a compact set in state-space, and is therefore dependent on the set A_T.

Given estimates of the coefficients $(c_j)_{j \in \mathcal{Z}^n}$, the size of A_T and the value of λ required to meet specifications on $\|\epsilon\|$ could be estimated. For example, an indication of the variation of ϵ with λ can be obtained from the L^2-norm of a function

$$\epsilon_0(\mathbf{x}) = \sum_{|p| > p_\lambda} \sum_{j \in J_p} c_j \phi_j(\mathbf{x}),$$

where J_p is the subset of J that indexes basis functions at scale p. If the candidate basis function set consists of radial wavelets and $f(\mathbf{x})$ is everywhere continuous on A_T, with Lipschitz exponent $\alpha \in (0,1)$, then using the bound derived in appendix 2.7, the magnitudes of coefficients $(c_j)_{j \in \mathcal{Z}^n}$ decay as $a_0^{(\alpha+1/2)np}$ at fine scales ($p \to -\infty$). Since the number of nodes $N(J_p)$ indexed by J_p is bounded by $N(J_0)$ for $p < 0$, and

by $a_0^{-np}N(J_0)$ for $p > 0$, $\|\epsilon_0\|$ can in this case be bounded by

$$\|\epsilon_0\|^2 \leq MN(J_0)\left[|f|_{\max}^2 |\tilde{\phi}|_{\max}^2 |A_T|^2 \left(\frac{a_0^{-np_\lambda}}{1-a_0^{-n}}\right) + C^2 m^2(\alpha)\left(\frac{a_0^{-2\alpha np_\lambda}}{1-a_0^{-2\alpha n}}\right)\right],$$

(where the parameters C and $m(\alpha)$ are defined in appendix 2.7). The maximum number of basis functions that can be incorporated in the expansion is in this case given by $N_0 = N(J_0)[p_\lambda + (a_0^{np_\lambda} - 1)/(1 - a_0^{-n})]$. By assigning costs $K_\epsilon(\|\epsilon_0\|)$ reflecting the desired convergence of the error measure, and $K_N(N_0)$ quantifying limits on the network's size imposed by finite computational resources, the optimal value of λ could be estimated from the value of p_λ that minimizes $K_\epsilon + K_N$.

2.6 Examples

Simulations of the space-frequency localized basis function networks described in this chapter are presented in this section. Figures 2-2 and 2-3 contrast the wavelet identification network's approximation to the function

$$f(x) = 0.25 + \mathrm{e}^{-x/10}\sin(0.16\pi x) - 0.1\cos(0.02\pi x - 0.5)$$

to that of a regularly sampled gaussian network. This function was chosen because of the significant variation in its local spatial bandwidth. The wavelet network employed "mexican hat" wavelets: $\phi(x) = (1 - x^2)\mathrm{e}^{-x^2/2}$, and both networks contained 20 nodes in the steady state. The errors in the wavelet network's reconstruction are approximately constant over $0 \leq x \leq 50$, showing its adaptation to the local spatial bandwidth of f. In contrast, the large errors in the function reconstructed by the gaussian network for $x < 10$ are due to the spatially uniform local bandwidth of its approximation. The wavelet network employed a threshold $\lambda = 0.069$, and bound $|f|_{\max} = 1.0$. The set A_T was defined by $0 \leq x \leq 50$, and the resulting limit on basis functions' scales was $p_\lambda = 13.4$. The maximum modulus of scale of basis functions introduced in the network expansion during training was $|p| = 5$.

Figure 2-4 shows the predictions of the mexican hat wavelet identification network

0.1 sec in the future of the velocity of the first order system

$$f(x) = -\text{sign}(x)x^2.$$

The network was trained over a single trajectory, using a 0.01 sec sample period. The large errors in the network's predictions at $x = 10$ are due to the use of initial conditions $\mathbf{x}(0) = 10$, $\hat{f}(\mathbf{x}(0)) = 0$. The distribution of the network's nodes on the space frequency lattice and the values of their associated coefficients are given in figure 2-5. The large coefficients of basis functions centred at increasingly fine spatial scales with increasing x is due to the increase in local spatial bandwidth of the approximated function as x increases.

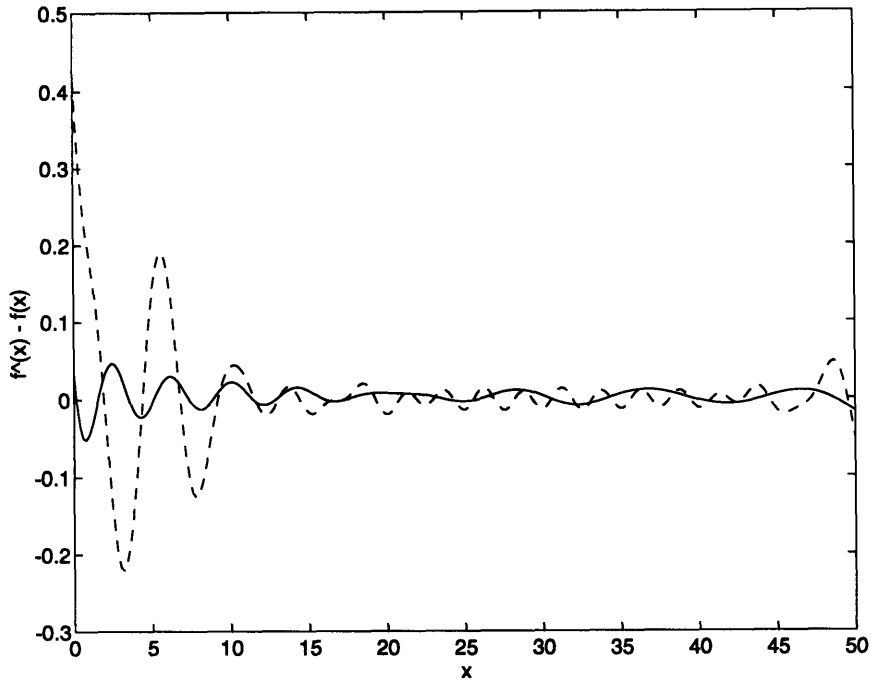

Figure 2-2: A comparison of the errors $\hat{f}(x) - f(x)$ in the recontructions of a wavelet network with 20 steady-state nodes and a gaussian network with 20 regularly spaced nodes. Solid line: wavelet network. Broken line: gaussian network.

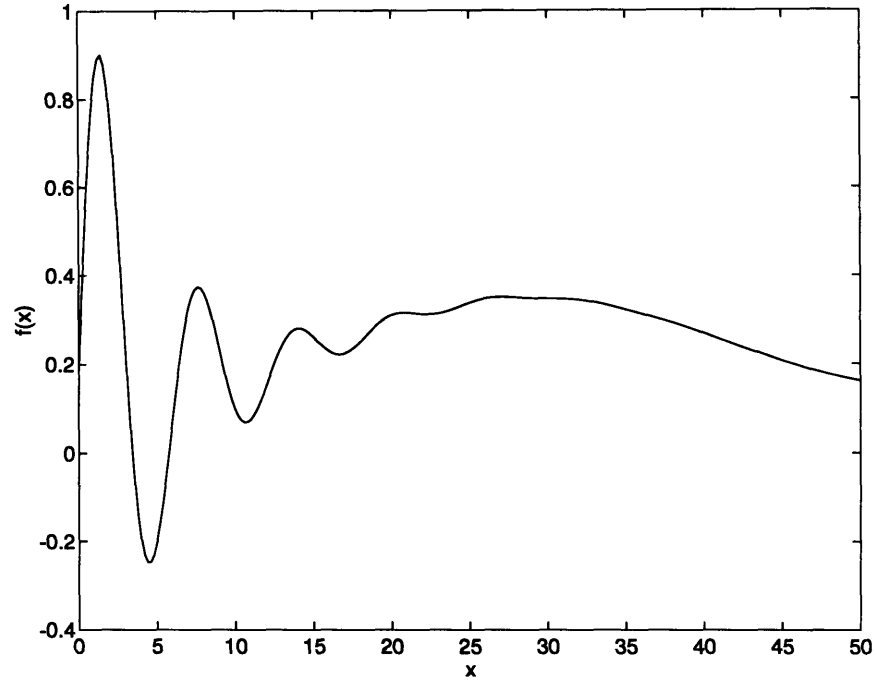

Figure 2-3: The function approximated in the example of figure 2-2, $f(x) = 0.25 + e^{-x/10}\sin(0.16\pi x) - 0.1\cos(0.02\pi x - 0.5)$

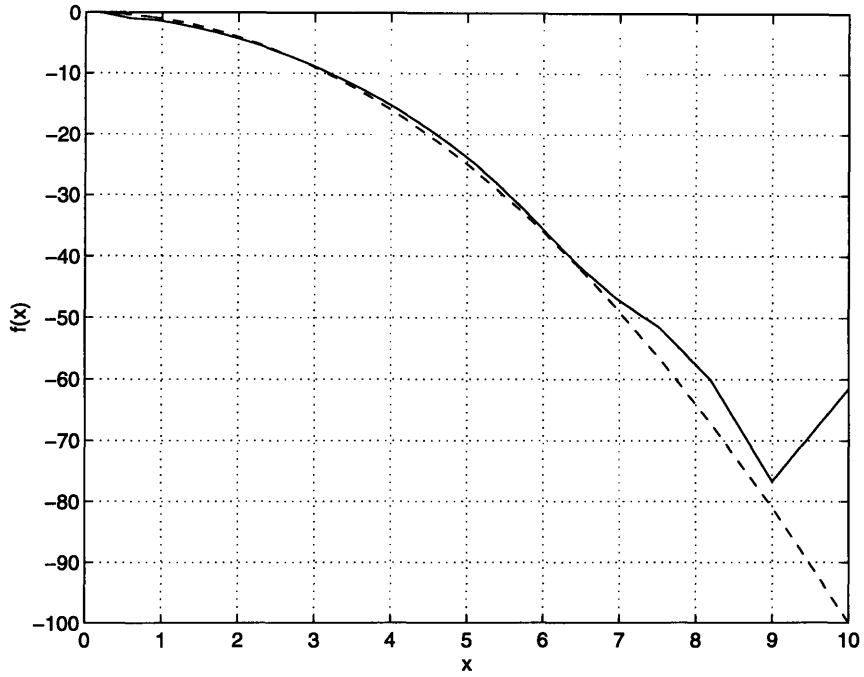

Figure 2-4: Wavelet identification network predictions of the value of $f(x(t+T))$ made at time t, where $f(x) = -\text{sign}(x)x^2$, and $T = 0.1$s. Solid line: the approximated function $f(x(t+T))$. Broken line: the the network's prediction $\hat{f}(x(t+T))$. The functions $\hat{f}(x(t+T))$ and $f(x(t+T))$ are plotted against $x(t+T)$.

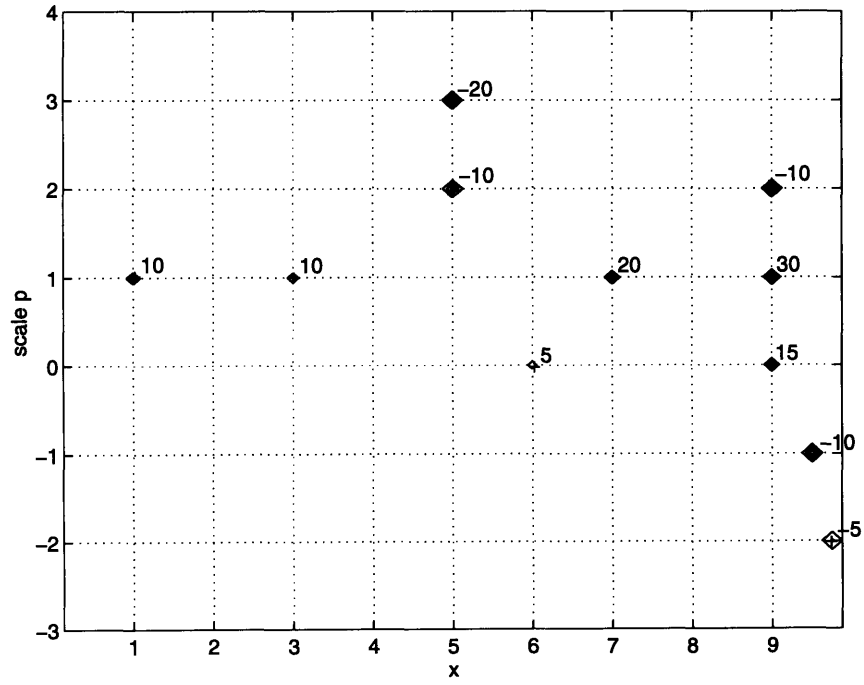

Figure 2-5: The steady state node distribution for the network of figure 2-4.

2.7 Appendix: Prior Bounds On Coefficients In Wavelet Expansions

Wavelet expansions of bounded functions with compact support The elements $\tilde{\phi}$ of the dual frame of a family of ridge or radial wavelets constituting a frame for $L^2(\mathcal{R}^n)$ can be expressed

$$\tilde{\phi}_{j(p,k)}(\mathbf{x}) = a_0^{-np/2}\tilde{\phi}_{j(0,k)}(a_0^{-np}\mathbf{x})$$

where the frame parameter a_0 satisfies $a_0 > 1$, and the dependence of the index j on the dilation and translation parameters p and k has been made explicit. From equation (2.2), the coefficients in an expansion over the family $(\phi)_{j\in\mathcal{Z}^n}$ giving exact reconstruction of $f_A(\mathbf{x})$ are bounded for all $j \in \mathcal{Z}^n$ by

$$\begin{aligned} |c_j| &\leq a_0^{np/2} \sup_{\mathbf{x}\in A_T} |f_A(\mathbf{x})| \int_{A_T} |\tilde{\phi}_{j(0,k)}(\mathbf{x})|\, dx \\ &\leq a_0^{np/2}\, |f|_{\max}\, |\tilde{\phi}|_{\max}\, |A_T|, \end{aligned}$$

where $|A_T| = \int_{A_T} dx$, $|f|_{\max} = \sup_{\mathbf{x}\in A_T} |f(\mathbf{x}) - c|$, and $|\tilde{\phi}|_{\max} = \sup_{\mathbf{x}\in A_T, k\in\mathcal{Z}^n} |\tilde{\phi}_{j(0,k)}(\mathbf{x})|$. Equation (2.2) also gives the alternative bound

$$\begin{aligned} |c_j| &\leq a_0^{-np/2} \sup_{\mathbf{x}\in A_T} |f_A(\mathbf{x})| \sup_{\mathbf{x}\in A_T} |\tilde{\phi}_{j(0,k)}(\mathbf{x})||A_T| \\ &\leq a_0^{-np/2}\, |f|_{\max}\, |\tilde{\phi}|_{\max}\, |A_T|, \end{aligned}$$

where $|f|_{\max}$, $|\tilde{\phi}|_{\max}$ and $|A_T|$ are as defined above.

The dual of a frame of separable n-dimensional wavelets has elements

$$\tilde{\phi}_{j(p,k)}(\mathbf{x}) = a_0^{-(p_1+\ldots+p_n)/2}\tilde{\varphi}_{j(0,k_1)}(a_0^{-p_1}x_1)\ldots\tilde{\varphi}_{j(0,k_n)}(a_0^{-p_n}x_n),$$

where $a_0 > 1$ and $x_1, \ldots, x_{n-1} \in \mathcal{R}$. It can be assumed without loss of generality that $p_1, \ldots, p_i > 0$ and $p_{i+1}, \ldots, p_n < 0$, so that the coefficients reconstructing $f_A(\mathbf{x})$ in an

expansion over $(\phi_j)_{j\in\mathcal{Z}^n}$ can be bounded for all j by

$$\begin{aligned}|c_j| &\leq a_0^{-(p_1+\ldots+p_i)/2+(p_{i+1}+\ldots+p_n)/2} \sup_{\mathbf{x}\in A_T}|f_A(\mathbf{x})| \sup_{x_1\in A_T}|\tilde{\varphi}_{j(0,k_1)}(x_1)|\ldots\sup_{x_i\in A_T}|\tilde{\varphi}_{j(0,k_i)}(x_i)| \\ &\quad \cdot\int_{A_T}|\tilde{\varphi}_{j(0,k_{i+1})}(x_{i+1})|\ldots|\tilde{\varphi}_{j(0,k_n)}(x_n)|\,dx \\ &\leq a_0^{-(|p_1|+\ldots+|p_n|)/2}|f|_{\max}|\tilde{\phi}|_{\max}|A_T|,\end{aligned}$$

where $|f|_{\max}$, $|\tilde{\phi}|_{\max}$ and $|A_T|$ are as defined above.

Coefficient bounds $|c_j|_{\max}$ for the three classes of multidimensional wavelet can therefore be expressed

$$|c_j|_{\max} = a_0^{-n|p|/2}\,|f|_{\max}\,|\tilde{\phi}|_{\max}\,|A_T|. \tag{2.23}$$

Radial wavelet expansions of compactly supported functions with Lipschitz regularity If $f(\mathbf{x})$ is uniformly Lipschitz α on the set A_T, with $\alpha\in(0,1)$, then

$$f(\mathbf{x}) = f(\mathbf{u}) + \epsilon_{\mathbf{u}}(\mathbf{x}),$$

where $|\epsilon_{\mathbf{u}}(\mathbf{x})| \leq C\|\mathbf{x}-\mathbf{u}\|^{\alpha}$ for all $\mathbf{x},\mathbf{u}\in A_T$, and some constant $C<\infty$. Since $f_A(\mathbf{x}) = m_T(\mathbf{x})(f(\mathbf{x})-c)$, and $m_T(\mathbf{x})$ is uniformly continuous with $|m_T(\mathbf{x})|\leq 1$ for all $\mathbf{x}\in A_T$, it follows that

$$f_A(\mathbf{x}) = f_A(\mathbf{u}) + \epsilon_{\mathbf{u}}(\mathbf{x}) \tag{2.24}$$

for all $\mathbf{x},\mathbf{u}\in A_T$. Since in addition, f_A is supported on A_T, it can be concluded that equation (2.24) holds for all $\mathbf{x},\ \mathbf{u}\ \in\mathcal{R}^n$, so that f_A is uniformly Lipschitz α on $\mathcal{R}^n$. The coefficients given by equation (2.2) that reconstruct f_A in an expansion over the family $(\phi_j)_{j\in\mathcal{Z}^n}$ can in this case be expressed

$$c_j = \int_{\mathcal{R}^n}(f_A(\mathbf{u})+\epsilon_{\mathbf{u}}(\mathbf{x}))\tilde{\phi}_j(\mathbf{x})\,dx.$$

Assuming that $\tilde{\phi}_j$ has at least one vanishing moment, so that $\int_{\mathcal{R}^n}\tilde{\phi}_j(\mathbf{x})\,dx = 0$

for all $j \in \mathcal{Z}^n$, it follows that

$$c_j = \int_{\mathcal{R}^n} \epsilon_{\mathbf{u}}(\mathbf{x}) \tilde{\phi}_j(\mathbf{x})\, dx.$$

This condition is invariably satisfied in practice, since the functions $\tilde{\phi}_j$ necessarily have one vanishing moment for all $j \in \mathcal{Z}^n$ in order that the set $(\tilde{\phi}_j)_{j \in \mathcal{Z}^n}$ constitutes a frame for $L^2(\mathcal{R}^n)$ [8]. The magnitudes of coefficients are therefore bounded by

$$|c_j| \leq C \int_{\mathbf{x} \in \mathcal{R}^n} \|\mathbf{x} - \mathbf{u}\|^\alpha |\tilde{\phi}_j(\mathbf{x})|\, dx$$

The elements of the dual of a frame of n-dimensional radial functions can be written $\tilde{\phi}_{j(p,k)}(\mathbf{x}) = a_0^{-np/2} \tilde{\phi}^k(a_0^{-np}\mathbf{x} - a_0^{-np}\mathbf{x}_k)$, for some set functions $(\tilde{\phi}^k)_{k \in \mathcal{Z}^n} \in L^2(\mathcal{R}^n)$, where $\mathbf{x}_k$ is the centre of spatial localization of $\tilde{\phi}_{j(0,k)}(\mathbf{x})$. This expression gives the bound

$$|c_j| \leq C a_0^{np/2} \int_{\mathcal{R}^n} \|a_0^{np}(\mathbf{x} - \mathbf{x}_k) - \mathbf{u}\|^\alpha |\tilde{\phi}^k(\mathbf{x})|\, dx.$$

Since f_A is uniformly Lipschitz on $\mathcal{R}^n$, a valid choice for $\mathbf{u}$ is $\mathbf{u} = -a_0^{np}\mathbf{x}_k$, so that

$$|c_j| \leq C a_0^{(\alpha+1/2)np} \int_{\mathcal{R}^n} \|\mathbf{x}\|^\alpha |\tilde{\phi}^k(\mathbf{x})|\, dx.$$

With the definition

$$m(\alpha) = \sup_{k \in \mathcal{Z}^n} \int_{\mathcal{R}^n} \|\mathbf{x}\|^\alpha |\tilde{\phi}^k(\mathbf{x})|\, dx, \tag{2.25}$$

the bound on coefficients can be written

$$|c_j| \leq a_0^{(\alpha+1/2)np} C m(\alpha). \tag{2.26}$$

2.8 Appendix: A Bound On the Network's Residual Error

In order to show that $|\epsilon(\mathbf{x},t)|$ is uniformly bounded on A, first note that $f(bfx)$ can be written for all $\mathbf{x} \in A$ as $f(\mathbf{x}) = \sum_{j \in Z^n} c_j \phi_j(\mathbf{x}) + c$, so that equation (2.10) can be re-written

$$\epsilon(\mathbf{x},t) = \sum_{j \in J_t^-} \hat{c}_j(t)\phi_j(\mathbf{x}) + \sum_{j \in J_t^+} c_j \phi_j(\mathbf{x}) - (f(\mathbf{x}) - c)$$

for all $\mathbf{x} \in A$. The function f and its mean value c on A_T are assumed to be finite on A. The boundedness of the remaining terms in the above expression can be shown by considering their components at each of the scales $0 \leq |p| < p_\lambda$. Let the component of ϵ that consists of linear superpositions of basis functions at scales of absolute magnitude $|p|$ be $\epsilon_{|p|}$. Then, since the scales p of all nodes in the sets J_t^- and J_t^+ are limited by $|p| < p_\lambda$, ϵ can be written

$$\epsilon(\mathbf{x},t) = \sum_{|p|=0}^{p_\lambda - 1} \epsilon_{|p|}(\mathbf{x},t) - (f(\mathbf{x}) - c)$$

for all $\mathbf{x} \in A$. Defining as $J_{|p|,t}^+$ and $J_{|p|,t}^-$ the subsets of J_t^+ and J_t^- respectively that contain the indices that correspond to scales of magnitude $|p|$, then $\epsilon_{|p|}$ is given by

$$\epsilon_{|p|}(\mathbf{x},t) = \sum_{j \in J_{|p|,t}^-} \hat{c}(t)\phi_j(\mathbf{x}) + \sum_{j \in J_{|p|,t}^+} c_j \phi(\mathbf{x}).$$

Since p_λ is finite, each sum in the above expression consists of a finite number of terms. Furthermore, the wavelet $\phi(\mathbf{x})$ is assumed to be bounded, and the coefficients in the in the sums are bounded by $|\hat{c}_j| \leq \lambda$ for all $j \in J_{|p|,t}^-$, and by $|c_j| \leq a_0^{-n|p|/2} |f|_{\max} |\tilde{\phi}|_{\max}$ for all $j \in J_{|p|,t}^+$. It follows that $\epsilon_{|p|}(\mathbf{x},t)$ is bounded for all $\mathbf{x} \in A$, and constants K^- and K^+ exist such that

$$|\epsilon_{|p|}(\mathbf{x},t)| \leq K^- a_0^{n|p|/2} + K^+$$

for $0 \leq |p| < p_\lambda$, and $\forall \mathbf{x} \in A$. Summing individual components, the following bound on ϵ is obtained

$$|\epsilon(\mathbf{x}, t)| \leq K^- \left(\frac{a_0^{np_\lambda/2} - 1}{1 - a_0^{-n/2}} \right) + K^+ p_\lambda + |(f(\mathbf{x}) - c)|. \tag{2.27}$$

It can therefore be seen that $\epsilon(\mathbf{x}, t)$ is uniformly bounded for all $\mathbf{x} \in A$, provided p_λ is finite.

Chapter 3

Candidate Basis Function Sets

The characteristics of families of functions admissible as the basis functions of the network described in chapter 2 are investigated in this chapter. The practical advantages of non-orthonormal frames over orthonormal bases are discussed. Candidate basis functions for separable, ridge and radial wavelet approximation schemes in arbitrarily high dimensional spaces are described. Using prior bounds on the temporal bandwidth of the dynamic system to be identified or controlled, we show how the redundancy of ridge and radial classes of candidate basis function set can be reduced without compromising the efficiency of approximation.

The stability properties of the adaptive control and identification schemes described in chapter 2.1 allow any number of additional parametric terms to be incorporated in the model $\hat{f}$ of equation (1.2) [37]. If a part of the system's dynamics is known to have a particular functional form, then the coefficients of these terms can therefore be estimated in parallel with a network approximating the dynamics of unknown structure, without affecting the global stability of the adaptive system. This property is exploited in the inclusion of the bias term $\hat{c}$ in the network (2.4).

The assumptions made in the derivations of structural and output weight adaptation laws that the candidate basis function set $(\phi_j)_{j\in\mathcal{Z}}$ has finite and strictly positive frame bounds require that the generating function $\phi(\mathbf{x})$ is well-localized simultaneously in space and spatial frequency [8, 7]. In fact, a very broad class of space-frequency localized families are admissible as candidate basis function sets for the

network. In practice however, the network's approximation capabilities are strongly dependent on the chosen family.

If the family of candidate functions form an orthonormal wavelet basis, the frame condition (2.1) specifies an energy conservation on the coefficients in the expansion. In this case, an estimate of the L^2-norm of the approximated function would allow the L^2-norm of the reconstruction error in the expansion to be estimated directly from its coefficients. Subject to conditions on the smoothness class of function approximated, this information would allow the threshold λ to be chosen on a statistical basis at the design stage [11, 12, 10]. Compactly supported orthonormal wavelet bases are not available in closed analytic form however, and are therefore not well suited to on-line control and estimation applications. Since non-orthonormal wavelets well localized in space and frequency can be obtained in simple closed-form expressions, candidate basis function sets constituting non-orthonormal frames are well-suited to the applications considered in this thesis.

3.1 Multidimensional Wavelet Basis Functions

In order to develop networks for estimation and control of systems of the form of equation (1.1) with $n > 1$, a candidate basis function set must be constructed from functions defined on $\mathcal{R}^n$ for arbitrary n. If the relative scales of spatial variation of f in different directions within $\mathcal{R}^n$ are initially unknown, basis functions $\phi_j(\mathbf{x})$ must be selected from a candidate set consisting of functions having independently varying dilation parameters associated with each spatial direction in order that an efficient model of the system can be developed.

This can be seen by considering, for example, the problem of approximating a bandlimited function $f(\mathbf{x})$ whose spatial bandwidth in a given direction is an order of magnitude greater than that in all other directions. A model incorporating a sufficiently large number of basis functions selected from a candidate set consisting of dilations and translations of a function having equal spatial bandwidths in each direction in $\mathcal{R}^n$ could approximate f to an arbitrarily high degree of accuracy, given

that this set constitutes a frame for the class of bandlimited functions. However, the Nyquist sampling theorem suggests that it would be possible to achieve the same accuracy of approximation using a model that employed an order of magnitude fewer basis functions, provided the candidate set contained translations and dilations of a function whose bandwidth in each direction matched that of f.

The problem of approximating the members of more general classes of unknown function can be approached using multidimensional wavelets. Several multidimensional generalizations of one-dimensional wavelet families are currently known. In the approach most commonly used in wavelet-based image compression techniques, the one-dimensional wavelet transform is applied separately in each of n orthogonal directions. The basis functions $\phi_j(\mathbf{x})$ of the associated reconstruction formula are therefore the products of n functions defined on single-dimensional spaces

$$\phi_j(\mathbf{x}) = \varphi_{j_1}(x_1)\varphi_{j_2}(x_2)\ldots\varphi_{j_n}(x),$$

where each j_i indexes a pair of integers (p_i, k_i), and $\varphi_{j_i}(x_i) = a_0^{-p_i/2}\varphi(a_0^{-p_i}x_i - k_i b_0)$. Owing to the separable nature of this decomposition, the set $(\phi_j)_{j\in\mathcal{Z}^n}$ constitutes a frame for $L^2(\mathcal{R}^n)$ if and only if $(\varphi_j)_{j\in\mathcal{Z}}$ forms a frame for $L^2(\mathcal{R})$. A necessary condition on the decay of ϕ in order that $(\varphi_j)_{j\in\mathcal{Z}}$ forms a frame for $L^2(\mathcal{R})$ is

$$2\pi\int_0^\infty |\mathcal{F}\varphi(\xi)|^2\,\frac{d\xi}{\xi} < \infty, \tag{3.1}$$

and sufficient conditions on the parameters a_0, b_0 are given in [8].

An alternative class of multidimensional wavelet is obtained from the rotations of template functions in addition to their dilates and translates in $\mathcal{R}^n$ [28, 24]. The reconstruction model corresponding to this scheme has basis functions

$$\phi_j(\mathbf{x}) = a_0^{-pn/2}\varphi_i(a_0^{-p}\mathbf{x} - \Lambda_\mathbf{k}\mathbf{b}_0). \tag{3.2}$$

Each function φ_i has the general form $\varphi_i(\mathbf{x}) = \varphi(\Lambda R_i^T\mathbf{x})$ where the set of matrices $(R_i)_{i=1}^L$ represent a regular sampling of the space of rotations on $\mathcal{R}^n$, and $\Lambda \in \mathcal{R}^{n\times n}$

is a diagonal matrix. The matrix $\Lambda_{\mathbf{k}} = \text{diag}\{k_1, \ldots, k_n\}$ determines the spatial centre of each wavelet, and $p,\ k_1, \ldots,\ k_n \in \mathcal{Z}$. Conditions on φ and on the parameters a_0, $\mathbf{b_0}$, and L to ensure that a frame for $L^2(\mathcal{R}^n)$ is obtained are given in [28, 29].

The class defined at (3.2) contains radial functions as a special case. By defining $\varphi_i(\mathbf{x}) = \varphi[r_{W_i}(\mathbf{x})]$, where $r_{W_i}(\mathbf{x}) = \|W_i\mathbf{x}\|_2$, and each member of the set $(W_i)_{i=1}^n$ is a positive definite symmetric matrix with eigenvalue matrix Λ and eigenvector matrix R_i, (3.2) specifies a family of wavelets radial in the measure r. In this case, the condition equivalent to (3.1) on the decay of φ necessary in order that $(\phi_j)_{j\in\mathcal{Z}^n}$ constitute a frame for $L^2(\mathcal{R}^n)$ is given by

$$(2\pi)^n \int_0^\infty |\mathcal{F}_\rho\varphi(\rho)|^2 \frac{d\rho}{\rho} < \infty, \tag{3.3}$$

(where $\mathcal{F}_\rho\varphi$ is the Fourier transform of φ in polar coordinates), [28, 25]. Moreover, with a set of radial measures $\{r_{W_i^1},\ r_{W_i^2}, \ldots, r_{W_i^M}\}$, where each matrix W_i^m has eigenvector matrix R_i and eigenvalue matrix Λ^m, equation (3.2) can be used to define a set of radial wavelet families $((\phi_j^m)_{j\in\mathcal{Z}^n})_{m=1}^M$ with independently varying scaling parameters in orthogonal directions. Since the Fourier transforms of the radial functions $\varphi_i^l(\mathbf{x})$ and $\varphi_i^m(\mathbf{x})$ are related by

$$\mathcal{F}_\rho\varphi_i^m(\rho) = \prod_{j=1}^n \frac{\sigma_j^l}{\sigma_j^m} \mathcal{F}_\rho\varphi_i^l(\rho),$$

where $\Lambda^m = \text{diag}\{\sigma_1^m, \ldots, \sigma_n^m\}$ and $\Lambda^l = \text{diag}\{\sigma_1^l, \ldots, \sigma_n^l\}$, the conditions on $a_0, \mathbf{b_0}$, and L to ensure that $(\phi_j^m)_{j\in\mathcal{Z}^n}$ constitute a frame for $L^2(\mathcal{R}^n)$ are identical to those imposed on the dilation, translation and rotation parameters of the family $(\phi_j^l)_{j\in\mathcal{Z}^n}$, although with $\Lambda^m \neq \Lambda^l$ the frame bounds associated with the two families will not be equal. It follows that frames for $L^2(\mathcal{R}^n)$ can be constructed using the collection of wavelet families $((\phi_j^m)_{j\in\mathcal{Z}^n})_{m=1}^M$, with each component family $(\phi_j^m)_{j\in\mathcal{Z}^n}$ employing the same set of parameters $a_0, \mathbf{b_0}$, and L.

Ridge functions are also included in the class defined at (3.2). If $y_i(\mathbf{x}) = \sigma\mathbf{e}_i^T\mathbf{x}$, where σ is a positive scalar and $\mathbf{e}_i = R_i\mathbf{e}$ for some unit vector $\mathbf{e} \in \mathcal{R}^n$, then with φ_i

of the form $\varphi_i = \varphi[y_i(\mathbf{x})]$, (3.2) defines a family of ridge wavelets. Although this class cannot form frames for $L^2(\mathcal{R}^n)$ since $\varphi \notin L^2(\mathcal{R}^n)$, the truncation of f to a compact set A_T described in section 2.1 allows the use of candidate basis function sets that constitute frames for $L^2(A_T)$. Analogously to the case of radial wavelets, frames can be constructed from collections of families of the form ϕ_j^m, each of which employs a different scale factor σ^m.

3.2 Radial And Ridge Wavelet Models Of Bandlimited Systems

Candidate basis function sets constructed from separable, radial and ridge wavelet families include many functions with the same dilation parameter in a particular direction at any given centre of localization in space. This section shows how a bound on the component of the gradient of the approximated function in the direction of the state-vector's derivative can be used to eliminate superfluous members of the candidate set as the network's model evolves. Owing to the variable direction in which this bound holds, it is more easily applied to radial and ridge wavelet families than separable basis function sets.

If the system to be identified or controlled is strictly bandlimited, so that the Fourier transform of $x^{(k)}(t)$ is supported on an interval $[-\Omega_k, \Omega_k]$, then from the structure of equation (1.1) it follows that $\Omega_k = \Omega$ is independent of k. Assuming $f \in L^\infty(\mathcal{R}^n)$, the application of an inequality due to S. Bernstein [25] yields a bound on the spatial gradient of $f(\mathbf{x})$

$$|\dot{\mathbf{x}}^T \nabla f(\mathbf{x})| \leq \Omega \sup_{\mathbf{x} \in \mathcal{R}^n} |f(\mathbf{x})|. \tag{3.4}$$

Before proceeding to show how this bound suggests an approach to determining the relative scalings of basis functions in a network reconstruction of f, the restrictions imposed by the assumptions that the system is bandlimited and that $f \in L^\infty(\mathcal{R}^n)$ will first be discussed.

The state variables of all physical systems encountered in practice vary on a finite time-scale. For the majority of such systems therefore, the energy contained in the Fourier transform $\mathcal{F}\mathbf{x}(\omega)$ of the state $\mathbf{x}(t)$ is sufficiently well-localized that a bandwidth Ω exists so that $|\mathcal{F}\mathbf{x}(\omega)| \approx 0$ for $\omega \notin [-\Omega, \Omega]$.

Additional limits on the effective bandwidth of the modeled system arise as a result of non-ideal sensor and actuator characteristics. If the available measurement of the state is a vector $\mathbf{x}^h(t) = [x_1^h(t) \ \ldots \ x_{n-1}^h(t)]^T$, with $x_k^h(t) = h_k(x^{(k)}(t))$, then owing to the low-pass characteristics of the functions h_k in general, the bandwidth of the best approximation to the system's state is limited by the maximum bandwidth $\Omega_{\max}^h$ of the functions $h_1, \ldots, h_{n-1}$.

Alternatively, the control input to the controlled system is in general of the form $g(u(t))$, and the assumption that $u(t) = g(u(t))$ is valid only if the bandwidth of u is less than some finite value Ω^g. Since the control signal (2.8) must be filtered in order to avoid exciting unmodeled actuator dynamics if $\Omega > \Omega^g$, it can be seen that the bandwidth of the signal $\hat{f}(\mathbf{x}(t))$ supplied by network model need not exceed Ω^g. The effective bandwidth of the approximated system is therefore the minimum of Ω, $\Omega_{\max}^h$, and Ω^g.

The second assumption that $f(\mathbf{x})$ is bounded on $\mathcal{R}^n$ can be weakened since in both control and identification applications, the function f_A modeled by the space-frequency localized component of the network is defined only on a compact subset A_T of state-space. It follows that for practical purposes, the requirement that $f \in L^\infty(\mathcal{R}^n)$ can be replaced by the condition that $|f|_{\max} = \sup_{\mathbf{x} \in A_T} |f_A(\mathbf{x})|$ exists, and the bound (3.4) can be replaced by

$$|\dot{\mathbf{x}}^T \nabla f(\mathbf{x})| \leq \Omega |f|_{\max}, \tag{3.5}$$

for all $\mathbf{x} \in A$ (where $f_A(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in A$ has been used).

In the absence of a control input, the time-scale of variations in the system's state is dependent on the rate of change of the forcing function f. This inequality therefore has the intuitive explanation that the scale of spatial variation of f along a

given trajectory necessarily increases at higher rates of motion through state space. The scalings of basis functions in a network model of the system should therefore vary according to the orientation of its trajectories and the rates of change of the state vector at each point in state space. Assuming the generating wavelet ϕ to have some degree of regularity, the asymptotic differential operator properties of the inner products $(< \, . \, , \phi_j >)_{j \in \mathcal{Z}^n}$ can be used to estimate the dependence of the coefficients in an exact reconstruction of a function on its gradient. This information allows the rates of decay across scale of the magnitudes of coefficients in a wavelet model to be approximately equalized at all points along a trajectory, so that basis functions are scaled according to the scale of local variation of f in the direction of motion in state space.

In order to extend the differential operator properties of the continuous wavelet transform on $L^2(\mathcal{R})$ [25, 22] to the multidimensional case, we assume that the generating wavelet in equation (3.2) has least one vanishing moment,

$$\int \varphi(x) \, dx = 0. \tag{3.6}$$

For radial wavelets, the integration is over $x \in \mathcal{R}^+$, and for ridge wavelets over $x \in \mathcal{R}$. It then follows that, if $\varphi \in L^2(\mathcal{R})$, there exists a function $\vartheta \in L^2(\mathcal{R})$ satisfying $\varphi(x) = \vartheta'(x)$ for all $x \in \mathcal{R}$ [25]. Defining continuously parameterized ridge and radial wavelets as $\phi_0^{a,\mathbf{b}}(\mathbf{x}) = a^{-n/2}\varphi(y[(\mathbf{x}-\mathbf{b})/a])$ and $\phi_1^{a,\mathbf{b}}(\mathbf{x}) = a^{-n/2}\varphi(r_W[(\mathbf{x}-\mathbf{b})/a])$ respectively, and denoting the gradient operator with respect to $\mathbf{b}$ as $\nabla_{\mathbf{b}}$, we have

$$\phi_0^{a,\mathbf{b}}(\mathbf{x}) = -\frac{a^{-n/2+1}}{\sigma_1 \mathbf{v}^T \mathbf{e}_1} \mathbf{v}^T \nabla_{\mathbf{b}} \left\{ \vartheta \left[y \left(\frac{\mathbf{x}-\mathbf{b}}{a} \right) \right] \right\}, \tag{3.7}$$

for all $\mathbf{v} \in \mathcal{R}^n$ such that $\mathbf{v}^T \mathbf{e}_1 \neq 0$, and

$$\phi_1^{a,\mathbf{b}}(\mathbf{x}) = a^{-n/2+1} \frac{s_{\vartheta'}[(\mathbf{b}-\mathbf{x})/a]}{\|A^T \nabla r_W(\mathbf{b}-\mathbf{x})\|_2} \left\| A^T \nabla_{\mathbf{b}} \vartheta \left[r_W \left(\frac{\mathbf{b}-\mathbf{x}}{a} \right) \right] \right\|_2, \tag{3.8}$$

where $s_{\vartheta'}(\mathbf{x}) = \text{sign}\{\vartheta'[r_W(\mathbf{x})]\}$ and $A \in \mathcal{R}^{n \times n}$ with $\text{rank}(A) = n$.

Using equations (3.7), (3.8), and the inequality (3.5), appendices 3.4.1 and 3.4.2

derive the following asymptotic bounds on the magnitude of the continuous wavelet transform $Wf_A(a, \mathbf{b})$ with respect to ridge and radial wavelets with one vanishing moment,

$$\lim_{a \to 0} \frac{|Wf_A(a, \mathbf{b})|}{a^{n/2+1}} \leq \frac{\Omega |f|_{\max}}{\|\dot{\mathbf{x}}(\mathbf{b})\|_2} \kappa(\sigma_1).$$

For ridge wavelets, the rotation parameter $\mathbf{e}_1$ is assumed to be aligned with $\dot{\mathbf{x}}(\mathbf{b})$, and κ is defined by

$$\kappa(\sigma_1) = \frac{1}{\sigma_1^2} \int_{A_T} |\vartheta(\mathbf{e}_1^T \mathbf{x})| \, dx.$$

In the case of radial wavelets, the scaling factor σ_1 is the eigenvalue corresponding to an eigenvector of W aligned with $\dot{\mathbf{x}}(\mathbf{b})$, and

$$\kappa(\sigma_1) = \left[\frac{1}{2\sigma_1^2} + \sum_{i=2}^{n} \frac{1}{\sigma_i^2} \right]^{1/2} \prod_{i=1}^{n} \frac{1}{\sigma_i} \int_{\mathcal{R}^n} |\vartheta[(\mathbf{x}^T \mathbf{x})]| \, dx.$$

To a first order approximation, the inner products of f with ϕ_j are therefore bounded by

$$| < f, \phi_j > | \leq \frac{\Omega |f|_{\max} a_0^{p(n/2+1)}}{\|\dot{\mathbf{x}}(\mathbf{x}_j)\|_2} \kappa(\sigma_1), \tag{3.9}$$

where $\mathbf{x}_j = a_0^p \Lambda_{\mathbf{k}} \mathbf{b}_0$ is the spatial centre of $\phi_j(\mathbf{x})$.

The coefficients in a wavelet model that are known to reconstruct f_A exactly are the inner products of f_A with the members $\tilde{\phi}_j$ of the dual frame, as discussed in chapter 2. However [8], §3.5 shows that the reconstruction formula $\sum_{j \in \mathcal{Z}^n} < f_A, \tilde{\phi}_j > \phi_j$ can be reasonably well approximated by $\sum_{j \in J} < f_A, \phi_j > \phi_j$ for a sufficiently large finite index set J. Bounds on the magnitudes of coefficients of fine-scale wavelets in a truncated expansion may therefore be estimated using (3.9).

If the scaling parameter σ_1 that corresponds to a basis function centred at $\mathbf{x}_j$ is chosen to satisfy

$$\frac{1}{\kappa(\sigma_1)} = \frac{\Omega |f|_{\max} a_0^{-p_\lambda(n/2+1)}}{\|\hat{\dot{\mathbf{x}}}(\mathbf{x}_j)\|_2}, \tag{3.10}$$

(where $\hat{\dot{\mathbf{x}}}(\mathbf{x}_j)$ is an estimate of the state vector's derivative incorporating the model $\hat{f}(\mathbf{x}_j)$), and its rotation parameter is selected so that $\mathbf{e}_1 = \hat{\dot{\mathbf{x}}}(\mathbf{x}_j) / \|\hat{\dot{\mathbf{x}}}(\mathbf{x}_j)\|_2$, then the approximate analysis given above shows that only coefficients at scales p approx-

imately equal to, or greater than $-p_\lambda$ will have steady-state values exceeding the threshold λ. By effectively adjusting the local values of the frame parameters a_0 and $\mathbf{b}_0$, this selection criterion therefore matches the scale of each basis function's spatial variation along the trajectory followed to an estimated upper bound on the local spatial frequency content of f. Moreover, for both radial and ridge classes of wavelet, $1/\kappa$ is $O(\sigma_1^2)$, so that equation (3.10) is in agreement with the intuition that basis functions' spatial frequency content should increase in regions that the system's state varies slowly, as (3.5) also suggests.

Owing to the dependence of the estimates $\hat{\dot{\mathbf{x}}}(\mathbf{x}_j)$ on the approximation $\hat{f}(\mathbf{x})$, errors in $\hat{f}(\mathbf{x})$ result in inaccurate estimates of the scalings and orientations of basis functions. The error in $\hat{\dot{\mathbf{x}}}(\mathbf{x}_j)$ is likely to increase with the distance of the point $\mathbf{x}_j$ from the followed trajectory, due to the larger errors in the network's reconstruction in these regions. The resulting errors in estimates of basis functions' orientations and scalings have little effect on the quality of the approximation however, since basis functions introduced at points distant from all points on the followed trajectory can only remain in the network expansion if their spatial supports are broad, whereas the orientations of coarse-scale basis functions are relatively unimportant.

3.3 Examples

The performance of control and identification networks employing the basis function selection algorithms developed above are illustrated in this section with simulations of two second-order systems. The generating wavelets used are two-dimensional extensions of the single-dimensional "mexican hat" functions: $\varphi(x) = (1 - x^2)\mathrm{e}^{-\alpha x^2}$, $\alpha \in \mathcal{R}$. Since

$$\int_{\mathcal{R}^+} \varphi(x)\, dx = 0,$$

the condition that ϕ have one vanishing moment is satisfied for both radial and ridge classes. However, the constraint on the decay of the radial wavelet's Fourier

transform (3.3) requires in addition that

$$\int_{\mathcal{R}^+} r\varphi(r)\, dr = 0.$$

This requirement cannot be met by the unmodified mexican hat wavelet since it necessitates that $\phi(r) = g''(r)$ for all $r \in [0, \infty)$ and some $g \in L^2(\mathcal{R}^+)$. Although $\phi(r)$ is the second derivative of the Gaussian for all $r > 0$, at $r = 0$ this condition cannot be met due to the non-zero value of the Gaussian. This technical difficulty can be overcome using the difference of two mexican hat generating functions: $\phi(r) = (1 - r^2)\mathrm{e}^{-\alpha r^2}(1 - \mathrm{e}^{-\beta r^2})$, for any $\alpha, \beta \in \mathcal{R}$. In practice a very large value for β can be used, so that the modified wavelet is effectively identical to the original mexican hat wavelet.

Figures 3-1, to 3-6 show the application of a ridge and a radial wavelet network to the identification of the stable second order system

$$f(\mathbf{x}) = \tanh(x) + \tanh(\dot{x}).$$

A bandwidth of 2 rad/s was assumed and a sample period of 0.01 sec was used. Both radial and ridge basis function networks were trained over the duration of a single trajectory, and the coefficient magnitude threshold λ was chosen in each so that 15 nodes were obtained in the steady state.

Figure 3-2 compares the errors in state predictions 0.1 second in the future for the two networks, and figures 3-3 and 3-4 give the corresponding predictions of acceleration. The radial wavelet network clearly has a higher degree of accuracy than the ridge wavelet network. This is a result of the lack of spatial localization of the ridge basis functions in the direction perpendicular to the estimate of the state vector's derivative.

Figure 3-5 shows the locations, orientations, and scalings of the radial wavelet network's basis functions as the loci of points $r_W(a_0^{-p}\mathbf{x} - \Lambda_\mathbf{k}\mathbf{b}_0) = 0.7$. Similarly figure 3-6 gives the lines that join the points that are closest to the wavelets' centres on the lines $y(a_0^{-p}\mathbf{x} - \Lambda_\mathbf{k}\mathbf{b}_0) = \pm 1.0$. The inaccuracies in the orientations of basis

functions with centres $\mathbf{x}_j$ at large distances from the followed trajectory are due to inaccuracies in the approximation $\hat{f}(\mathbf{x})$ in these regions, which results in errors in the estimates $\hat{\dot{\mathbf{x}}}(\mathbf{x}_j)$.

Figures 3-7 to 3-10 give the results of simulating the control of the system

$$f(\mathbf{x}) = \cos(\pi x/2)\cos(\pi \dot{x}/2)\exp\{-[(x+2)^2 + (\dot{x}-2)^2]/2\} + 0.1\dot{x}cos(\pi x/5),$$

using a radial wavelet network and a p.d. controller. This system was selected for the high variation in the local spatial frequency content of the function f. The bandwidth was taken to be 10 rad/s, and a sample period of 0.01 sec was used. Figure 3-8 shows the cyclic desired trajectory to be followed, and the associated tracking errors and control inputs for both controllers are given in figures 3-10 and 3-9 respectively. It can be seen that the feed-forward provided by the network allows significant improvements in the controller's performance over the non-adaptive p.d. case. The wavelet network's steady state node distribution is shown in figure 3-11 as the loci $r_W(a_0^{-p}\mathbf{x} - \Lambda_\mathbf{k}\mathbf{b}_0) = 0.7$.

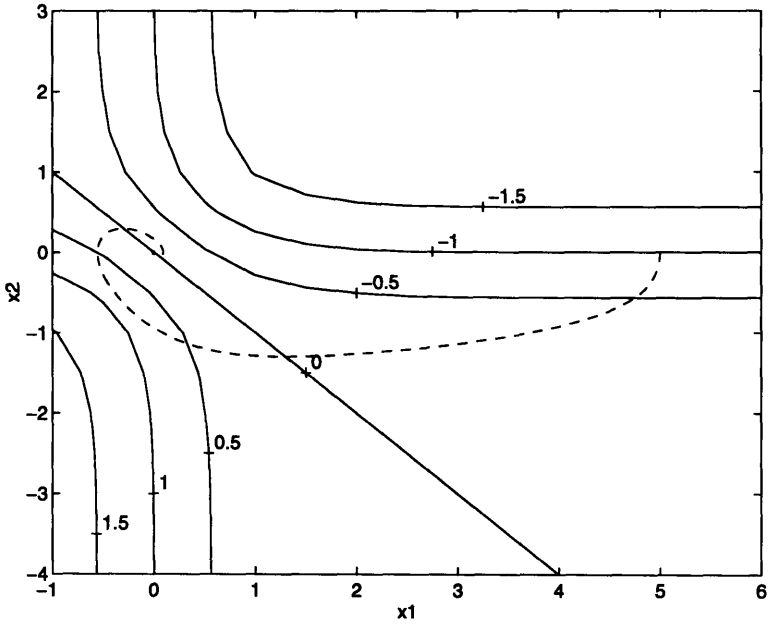

Figure 3-1: Contours of the approximated function $f(\mathbf{x}) = \tanh(x) + \tanh(\dot{x})$, and the trajectory followed in the examples of figures 3-2 to 3.6.

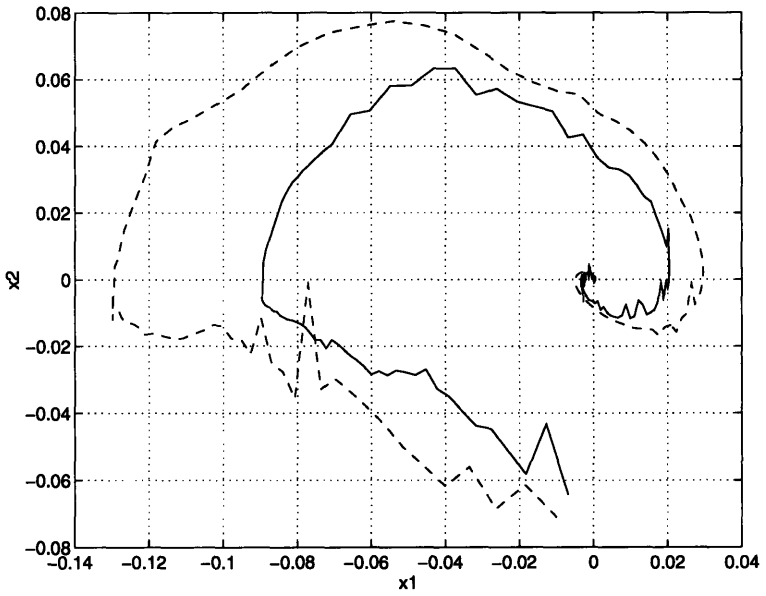

Figure 3-2: A comparison of errors in predictions $\hat{\mathbf{x}}(t+T)$ of the state of the second-order system $f(\mathbf{x}) = \tanh(x) + \tanh(\dot{x})$ made at time t, where $T = 0.1$s, using radial and ridge wavelet identification networks. Solid line: radial wavelet network prediction errors. Broken line: ridge wavelet prediction errors.

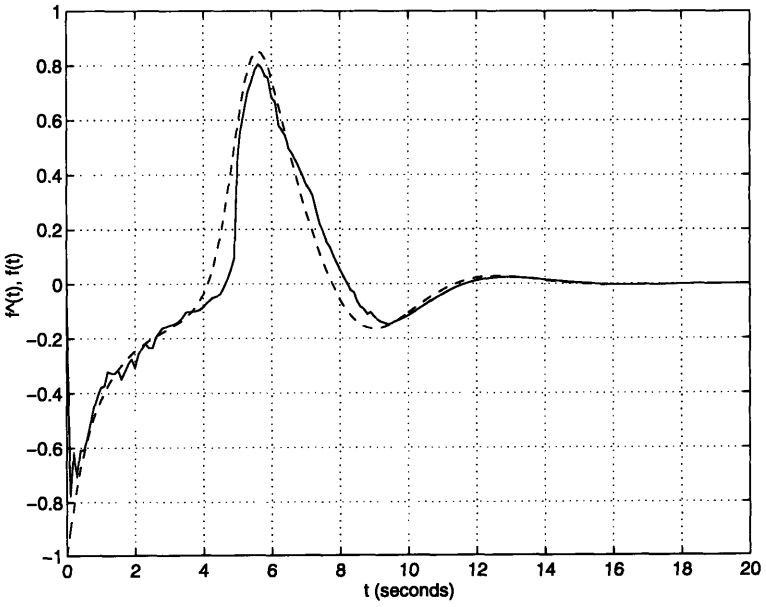

Figure 3-3: Predictions $\hat{f}(\mathbf{x}(t+T))$ of the acceleration of the second-order system $f(\mathbf{x}) = \tanh(x) + \tanh(\dot{x})$ made at time t, where $T = 0.1$s, using a radial wavelet identification network. Solid line: radial wavelet network predictions $\hat{f}(\mathbf{x}(t+T))$. Broken line: actual accelerations $f(\mathbf{x}(t+T))$.

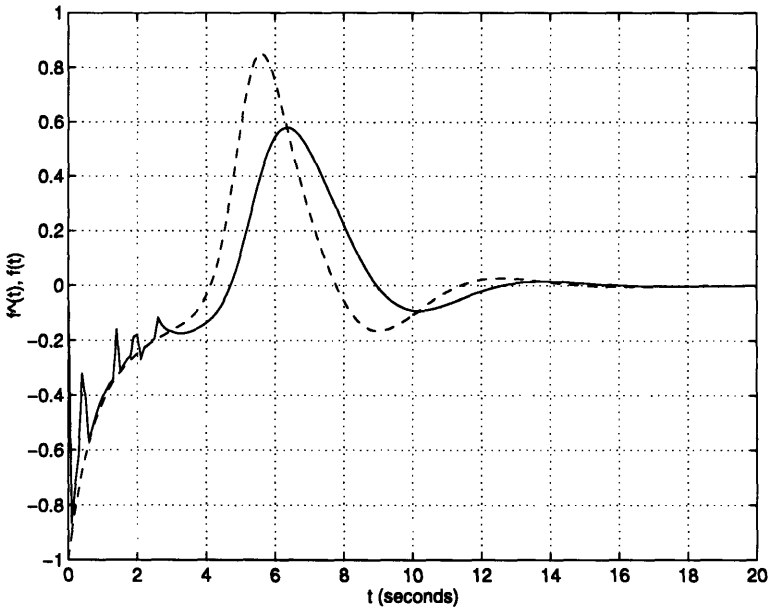

Figure 3-4: Predictions $\hat{f}(\mathbf{x}(t+T))$ of the acceleration of the second-order system $f(\mathbf{x}) = \tanh(x) + \tanh(\dot{x})$ made at time t, where $T = 0.1$s, using a ridge wavelet identification network. Solid line: ridge wavelet network predictions $\hat{f}(\mathbf{x}(t+T))$. Broken line: actual accelerations $f(\mathbf{x}(t+T))$.

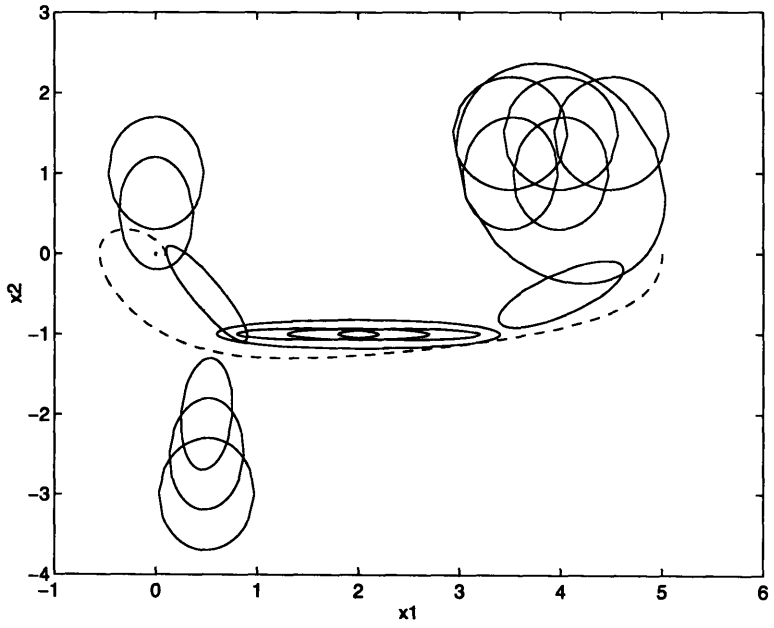

Figure 3-5: Orientations and scalings of the steady state nodes of the radial wavelet identification network.

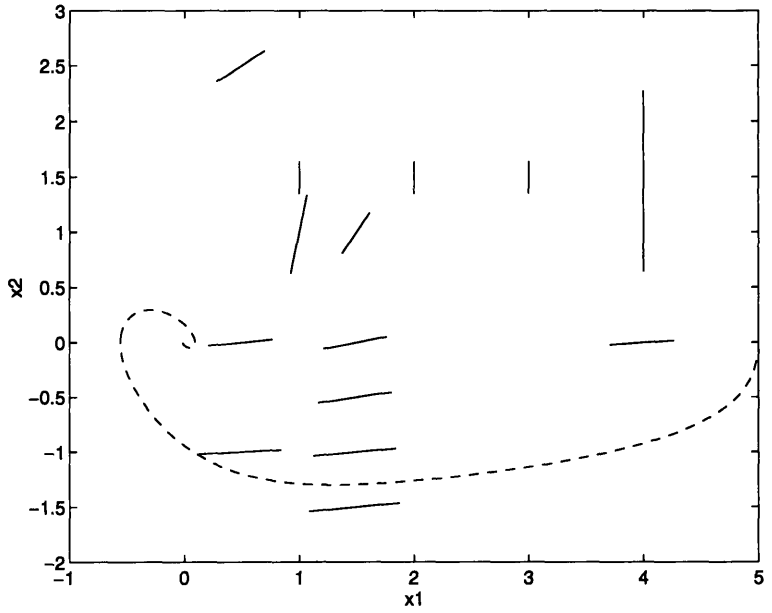

Figure 3-6: Orientations and scalings of the steady state nodes of the ridge wavelet identification network.

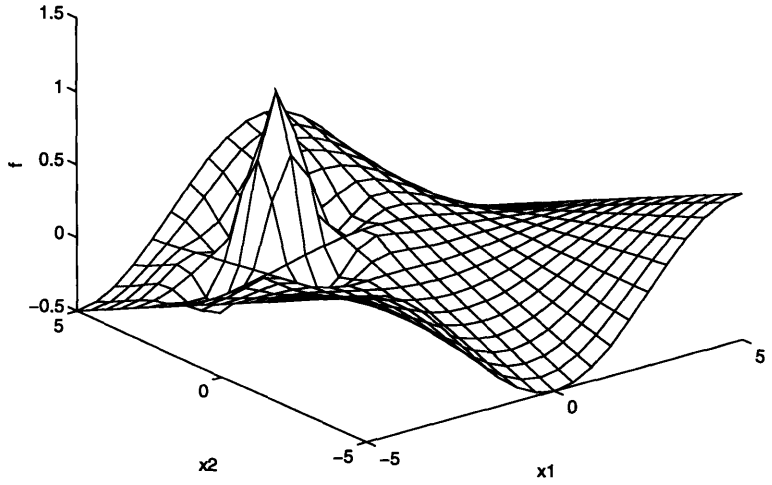

Figure 3-7: The controlled system in the examples of figures 3-8 to 3-11: $f(\mathbf{x}) = \cos(\pi x/2)\cos(\pi \dot{x}/2)\exp\{-[(x+2)^2 + (\dot{x}-2)^2]/2\} + 0.1\dot{x}cos(\pi x/5)$.

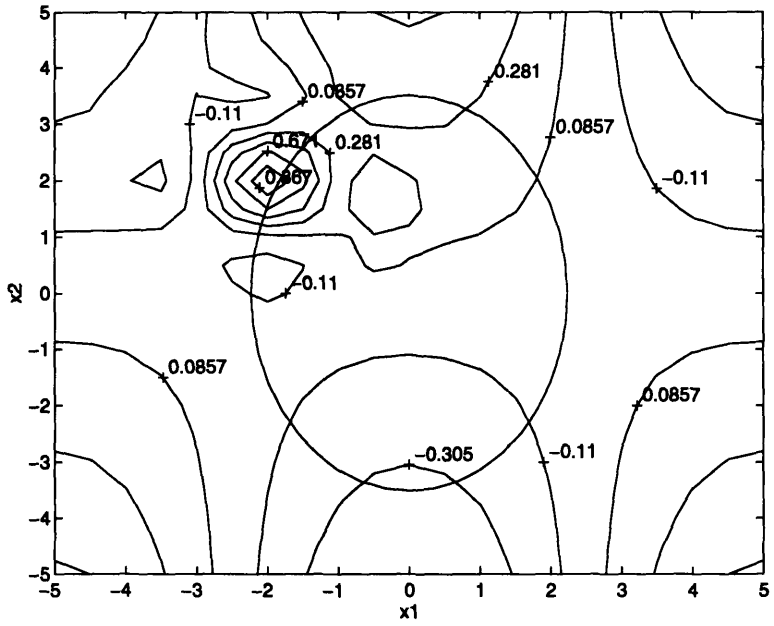

Figure 3-8: A contour map of the controlled system and the desired trajectory.

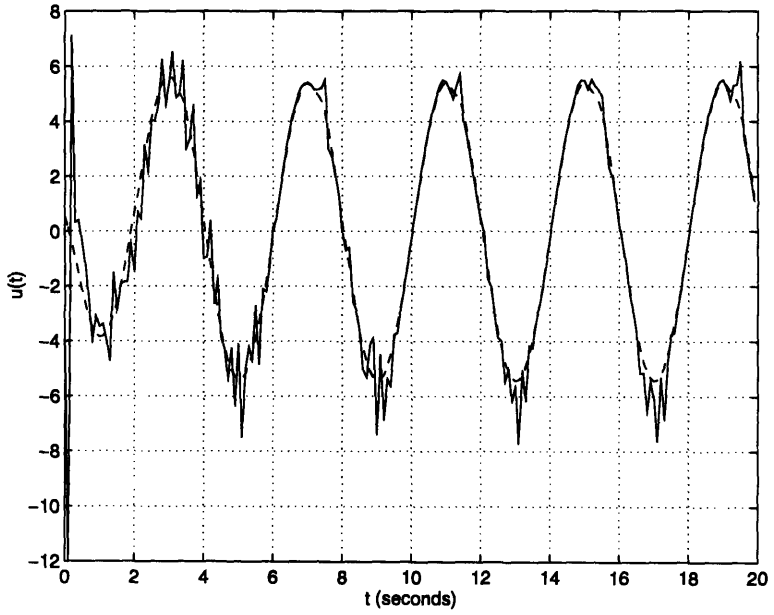

Figure 3-9: The control signals generated by radial wavelet network and p.d. controllers. Solid line: radial wavelet network controller. Broken line: p.d. comtroller.

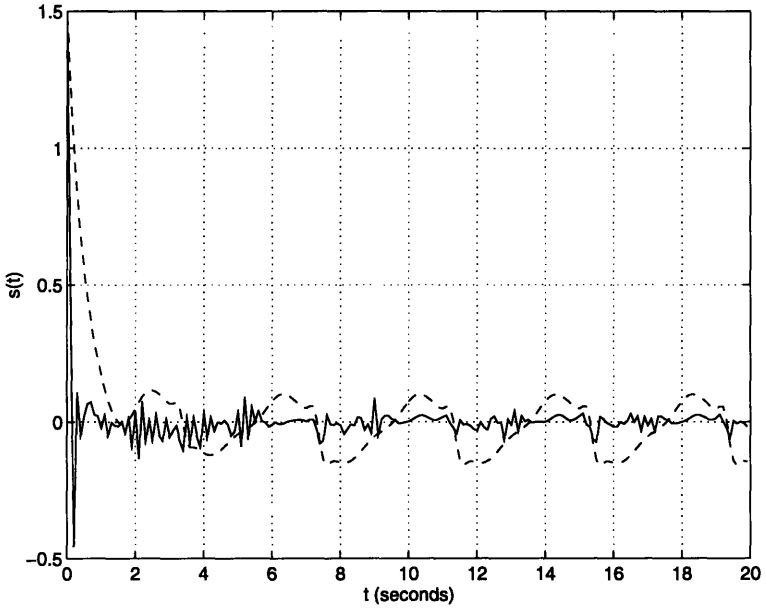

Figure 3-10: A comparison of the tracking errors for radial wavelet network and p.d. controllers. Solid line: radial wavelet network controller tracking errors. Broken line: p.d. controller tracking errors

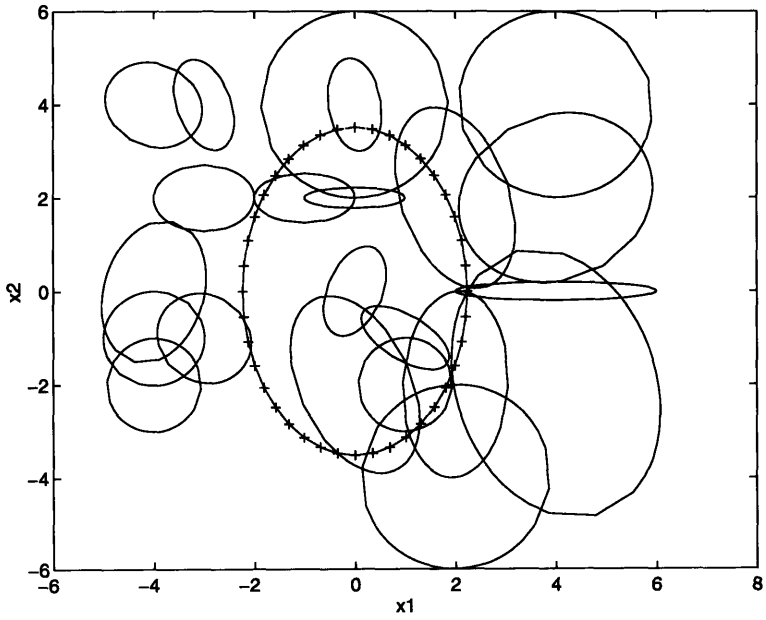

Figure 3-11: Steady state node orientations and scalings for the radial wavelet control network.

3.4 Appendix: Rates Of Decay Of Ridge And Radial Wavelet Transforms

3.4.1 Ridge Wavelets

For $f \in L^2(A_T)$, the continuous wavelet transform with respect to the ridge wavelet $\phi \in L^2(A_T)$ is defined

$$Wf(a, \mathbf{b}) = \int_{A_T} f(\mathbf{x}) \phi^{a,\mathbf{b}}(\mathbf{x})\, dx.$$

If $\varphi(y)$ has one vanishing moment, then equation (3.7) allows this expression to be re-written in terms of $\vartheta(y)$ as follows

$$\begin{aligned} Wf(a, \mathbf{b}) &= -a^{-n/2+1} \int_{A_T} \frac{f(\mathbf{x})}{\sigma_1 \mathbf{v}^T \mathbf{e}_1} \mathbf{v}^T \nabla_{\mathbf{b}} \left\{ \vartheta \left[y \left(\frac{\mathbf{x} - \mathbf{b}}{a} \right) \right] \right\} dx \\ &= \frac{-a^{n/2+1}}{\sigma_1 \mathbf{v}^T \mathbf{e}_1} \int_{A_T} \mathbf{v}^T \nabla_{\mathbf{b}} f(\mathbf{b} + a\mathbf{x})\, \vartheta[y(\mathbf{x})]\, dx. \end{aligned}$$

(The assumption that the controlled or identified system is strictly bandlimited ensures the existence of ∇f). The asymptotic decay of $|Wf(a, \mathbf{b})|$ at fine scales is therefore given by

$$\lim_{a \to 0} \frac{|Wf(a, \mathbf{b})|}{a^{n/2+1}} = \left| \frac{\mathbf{v}^T \nabla f(\mathbf{b})}{\sigma_1 \mathbf{v}^T \mathbf{e}_1} \right| \left| \int_{A_T} \vartheta[y(\mathbf{x})]\, dx \right|,$$

and the choices $\mathbf{e}_1 = \dot{\mathbf{x}}(\mathbf{b}) / \|\dot{\mathbf{x}}(\mathbf{b})\|_2$, and $\mathbf{v} = \dot{\mathbf{x}}(\mathbf{b})$ yield the bound

$$\lim_{a \to 0} \frac{|Wf(a, \mathbf{b})|}{a^{n/2+1}} \leq \frac{|\dot{\mathbf{x}}(\mathbf{b})^T \nabla f(\mathbf{b})|}{\sigma_1 \|\dot{\mathbf{x}}(\mathbf{b})\|_2} \int_{A_T} |\vartheta(\mathbf{e}_1^T \mathbf{x})|\, dx.$$

3.4.2 Radial Wavelets

With respect to the wavelet $\phi \in L^2(\mathcal{R}^n)$, radial in the metric $r_W(\mathbf{x})$, the continuous wavelet transform of a function $f \in L^2(\mathcal{R}^n)$ can be expressed

$$Wf(a, \mathbf{b}) = \int_{\mathcal{R}^n} f(\mathbf{x}) \phi^{a,\mathbf{b}}(\mathbf{x})\, dx.$$

If the analyzing wavelet has one vanishing moment, then it follows from equation (3.8) that for any full-rank matrix $A \in \mathcal{R}^{n\times n}$, $Wf(a,\mathbf{b})$ is given by

$$\begin{aligned} Wf(a,\mathbf{b}) &= a^{-n/2+1}\int_{\mathcal{R}^n} \frac{f(\mathbf{x})s_{\vartheta'}[(\mathbf{b}-\mathbf{x})/a]}{\|A^T\nabla r_W(\mathbf{b}-\mathbf{x})\|_2}\left\|A^T\nabla_{\mathbf{b}}\vartheta\left[r_W\left(\frac{\mathbf{b}-\mathbf{x}}{a}\right)\right]\right\|_2 dx \\ &= a^{n/2+1}\int_{\mathcal{R}^n}\frac{\|A^T\nabla_{\mathbf{b}}f(\mathbf{b}-a\mathbf{x})\|_2}{\|A^T\nabla r_W(\mathbf{x})\|_2}s_{\vartheta'}(\mathbf{x})\vartheta[r_W(\mathbf{x})]dx. \end{aligned}$$

The asymptotic decay of $Wf(a,\mathbf{b})$ as $a\to 0$ is therefore given by

$$\lim_{a\to 0}\frac{|Wf(a,\mathbf{b})|}{a^{n/2+1}} = \|A^T\nabla f(\mathbf{b})\|_2\left|\int_{\mathcal{R}^n}\frac{s_{\vartheta'}(\mathbf{x})\vartheta[r_W(\mathbf{x})]}{\|A^T\nabla r_W(\mathbf{x})\|_2}dx\right|. \tag{3.11}$$

Let the eigenvalues and eigenvectors of W be $(\sigma_i,\mathbf{e}_i)_{i=1}^n$. Assuming that $\mathbf{v}$ is not perpendicular to $\mathbf{e}_1$, a valid choice for the matrix A is $A=[\mathbf{a}_1\ k\mathbf{a}_2\ \ldots\ k\mathbf{a}_n]$, with $\mathbf{a}_1=\mathbf{e}_1$ and $\mathbf{a}_i^T\mathbf{v}=0$, $\mathbf{a}_i^T\mathbf{a}_j=\delta_{i,j}$ for $i=2,\ldots,n$, where $\mathbf{v}=\nabla f(\mathbf{b})/\|\nabla f(\mathbf{b})\|_2$ and $k\in\mathcal{R}$. This choice gives

$$\frac{\|A^T\nabla f(\mathbf{b})\|_2}{\|A^T\nabla r_W(\mathbf{x})\|_2} = |\mathbf{e}_1^T\nabla f(\mathbf{b})|\left[\frac{\|W\mathbf{x}\|_2^2}{k^2\|W^2\mathbf{x}\|_2^2-k^2(\mathbf{v}^TW^2\mathbf{x})^2+(\mathbf{e}_1^TW^2\mathbf{x})^2}\right]^{1/2}.$$

The assumption that $\mathbf{e}_1^T\nabla f(\mathbf{b})\neq 0$ ensures that $k=[1-\max_i(\mathbf{e}_i^T\mathbf{v})^2]^{-1/2}$ is well-defined, and the above expression can be bounded as follows

$$\frac{\|A^T\nabla f(\mathbf{b})\|_2}{\|A^T\nabla r_W(\mathbf{x})\|_2} = |\mathbf{e}_1^T\nabla f(\mathbf{b})|\left[\frac{\|W\mathbf{x}\|_2^2}{\|W^2\mathbf{x}\|_2^2+(\mathbf{e}_1^TW^2\mathbf{x})^2-k^2[(\mathbf{v}^TW^2\mathbf{x})^2-\max_i(\mathbf{e}_i^T\mathbf{v})^2\|W^2\mathbf{x}\|_2^2]}\right]^{1/2}.$$

Using the relationships

$$(\mathbf{v}^TW^2\mathbf{x})^2-\max_i(\mathbf{e}_i^T\mathbf{v})^2\|W^2\mathbf{x}\|_2^2 \le \frac{\max_i(\mathbf{e}_i^T\mathbf{v})^2}{\|\mathbf{x}\|_2^2}[(\mathbf{x}^TW^2\mathbf{x})^2-\|\mathbf{x}\|_2^2\,\|W^2\mathbf{x}\|_2^2]\le 0,$$

the following upper bound is obtained

$$\frac{\|A^T\nabla f(\mathbf{b})\|_2}{\|A^T\nabla r_W(\mathbf{x})\|_2} \le |\mathbf{e}_1^T\nabla f(\mathbf{b})|\left[\frac{\|W\mathbf{x}\|_2^2}{\|W^2\mathbf{x}\|_2^2+(\mathbf{e}_1^TW^2\mathbf{x})^2}\right]^{1/2}$$

$$\leq \quad |\mathbf{e}_1^T \nabla f(\mathbf{b})| \left[\frac{1}{2\sigma_1^2} + \sum_{i=2}^{n} \frac{1}{\sigma_i^2} \right]^{1/2}.$$

The decay of the wavelet transform (3.11) is therefore subject to the bound

$$\begin{aligned} \lim_{a \to 0} \frac{|Wf(a,\mathbf{b})|}{a^{n/2+1}} &\leq |\mathbf{e}_1^T \nabla f(\mathbf{b})| \left[\frac{1}{2\sigma_1^2} + \sum_{i=2}^{n} \frac{1}{\sigma_i^2} \right]^{1/2} \left| \int_{\mathcal{R}^n} \vartheta[r_W(\mathbf{x})] s_{\vartheta'}(\mathbf{x})\, dx \right| \\ &\leq |\mathbf{e}_1^T \nabla f(\mathbf{b})| \left[\frac{1}{2\sigma_1^2} + \sum_{i=2}^{n} \frac{1}{\sigma_i^2} \right]^{1/2} \prod_{i=1}^{n} \frac{1}{\sigma_i} \int_{\mathcal{R}^n} |\vartheta[(\mathbf{x}^T \mathbf{x})]|\, dx. \end{aligned}$$

Provided $\dot{\mathbf{x}}(\mathbf{b})^T \nabla f(\mathbf{b}) \neq 0$, by defining $\mathbf{e}_1 = \dot{\mathbf{x}}(\mathbf{b}) / \|\dot{\mathbf{x}}(\mathbf{b})\|_2$, we have

$$\lim_{a \to 0} \frac{|Wf(a,\mathbf{b})|}{a^{n/2+1}} \leq \frac{|\dot{\mathbf{x}}(\mathbf{b})^T \nabla f(\mathbf{b})|}{\|\dot{\mathbf{x}}(\mathbf{b})\|_2} \left[\frac{1}{2\sigma_1^2} + \sum_{i=2}^{n} \frac{1}{\sigma_i^2} \right]^{1/2} \prod_{i=1}^{n} \frac{1}{\sigma_i} \int_{\mathcal{R}^n} |\vartheta[(\mathbf{x}^T \mathbf{x})]|\, dx.$$

The assumption that the derivative of the state vector is not perpendicular to the the gradient of f is necessary as a result of the lack of information on the local variation of f in the radial wavelet transform in the direction of $\dot{\mathbf{x}}$ in this case. In practice this condition does not cause a significant restriction of the applicability of the bound derived above since the bound (3.5) on the gradient of f is inaccurate in this limiting case.

Chapter 4

Truncated Wavelet Expansions As Regularization Approximations

This chapter investigates the optimality of the approximations generated by the network developed in chapter 2. If the candidate basis function set constitutes an orthonormal basis, the criterion for selecting nodes for inclusion in the network on the basis of thresholding rules applied to coefficient estimates approximately minimizes the L^2-norm of the residual approximation error. In the more general case of the candidate basis function set constituting a non-orthonormal frame however, the same laws for structural adaptation can be shown to minimize approximately only a bound on the residual error.

For a class of non-orthonormal wavelet basis functions, it is shown that thresholding rules are optimal with respect to a regularization cost functional incorporating a smoothness measure that serves to quantify the degree to which the network's basis function set is matched to the wavelet transform of the approximated function. Since the regions of localization of the approximated function's wavelet transform on the space-frequency domain can be estimated from the magnitudes of coefficient estimates, the network's structural adaptation laws are approximately optimal for this class of wavelet basis function.

4.1 Optimal Reconstruction From Partial Data

An estimate of the dependence of the residual error in the network's approximation on the basis function set incorporated in its expansion would allow the set having least estimated residual error to be selected at each instant during training. Since the network's mean square reconstruction error is guaranteed to converge asymptotically to within a bound (2.22) dependent on the approximation capability of future basis function sets, an approximately optimal node distribution would then be obtained in the limit of convergence of adaptation. In the case of a candidate basis function set constituting an orthonormal basis, an estimate of this kind can be obtained from the network's estimated coefficients.

If basis functions are selected from the members of an orthonormal family, the coefficients $(c_j)_{j\in\mathcal{Z}}$ reconstructing the approximated function are unique, and the basis function set that minimizes the L^2-norm of the residual approximation error ϵ can be deduced directly from the set $(c_j)_{j\in\mathcal{Z}}$. Extending the definition (2.10) of ϵ to the whole of $\mathcal{R}^n$ as described in section 2.5, the energy conservation satisfied by the coefficients in an expansion over an orthonormal basis set allows $\|\epsilon\|^2$ to be written

$$\|\epsilon(\mathbf{x},t)\|^2 = \sum_{j\in J_t^-} \tilde{c}_j^2(t) + \sum_{j\in J_t^0} c_j^2 + \sum_{j\notin J} c_j^2.$$

Owing to the truncation of the network's expansion to the basis functions with centres lying on a compact subset of state-space, only nodes indexed by J can be selected for inclusion in the network. The value of $\|\epsilon\|$ is therefore minimized if the coefficients of least magnitude in the set $(c_j)_{j\in J_t^0}$ are neglected in the expansion, so that the index set of basis functions retained in the expansion satisfies $J_t^+ = \{j \in J_t^+ \cup J_t^0; |c_j| > \lambda\}$ for some threshold λ controlling the network's approximation capability. The basis function selection algorithm described in section 2.1 approximates this condition by thresholding instead the coefficient estimates.

In the case of a non-orthonormal basis function set however, the intuitively appealing strategy of constructing a model using wavelets having greatest correlation

with the approximated function cannot be justified by frame theory alone. This is a consequence of the linear dependence of basis functions, which has the effect that the L^2-norm of the residual approximation error cannot be expressed in terms of the sum of squared coefficient errors in the set J_t^- and the squares of the coefficients neglected in the expansion. An expression of this form, obtained from the bound on the norm of the adjoint frame operator, gives only a bound on the value of $\|\epsilon\|$,

$$\|\epsilon(\mathbf{x},t)\|^2 \leq M\left[\sum_{j\in J_t^-}\tilde{c}_j^2(t)+\sum_{j\in J_t^0}c_j^2+\sum_{j\notin J}c_j^2\right],$$

where M is the upper frame bound associated with the candidate basis function set $(\phi_j)_{j\in\mathcal{Z}}$. In this case, it cannot therefore be concluded that a basis function selection algorithm retaining only the coefficients greatest in magnitude in the expansion approximately minimizes $\|\epsilon\|$.

The optimality properties of a model employing the basis functions having greatest correlation with the approximated function can however be determined for a class of wavelet using regularization principles. Regularization has been widely used in the context of multivariate approximation to find unique solutions to the ill-posed problem of optimal function reconstruction from sparse data [13, 39, 17, 16].

The essence of the approach is to construct an approximation $\hat{f}$, making use of prior knowledge of the smoothness of the function to be approximated f_A in addition to a set of data $(x_i, f_A(x_i))_{i=1}^N$, obtained by random sampling f_A in the presence of noise. The smoothness information allows assessment of the quality of fit of $\hat{f}$ using a global measure, not simply the reconstruction errors at sample points alone. Subject to conditions on the form of the smoothness prior, this additional information provides a sufficient degree of constraint that the approximant $\hat{f}$ can be obtained as the solution to an optimization problem.

The optimal approximant is formulated as the minimizing argument of a cost functional $P[\hat{f}]$:

$$P[\hat{f}]=\sum_{i=1}^{N}[\hat{f}(x_i)-f_A(x_i)]^2+\lambda_0 K[\hat{f}].$$

The functional $K[\hat{f}]$ reflects the prior smoothness assumption by penalizing deviations from the anticipated degree of regularity of f_A, and the parameter $\lambda_0 \in \mathcal{R}^+$ reflects the trade-off between the error of approximation and satisfaction of the smoothness constraint. Provided K is chosen suitably, the solution $\hat{f}$ can be obtained via a variational argument, as we now detail further.

4.2 The Regularization Approximation Obtained Using A Smoothness Measure Based On The Wavelet Transform

The functional form of an approximation obtained from regularization principles is dictated by the smoothness functional employed. A common choice is the L^2-norm of $D^k\hat{f}$, where D^k is a linear differential operator of order k, leading to kth order polynomial spline approximations [14]. Generalizations of this type of smoothness measure having the form (in the single-dimensional case)

$$K[\hat{f}] = \int \frac{|\mathcal{F}\hat{f}(\xi)|^2}{\mathcal{F}g(\xi)}\, d\xi, \tag{4.1}$$

where $\mathcal{F}g$ denotes the Fourier transform of any conditionally positive definite function g, have also been shown to be valid choices for K [20]. In particular, the functional forms of a variety of radial basis function networks have been derived using smoothness measures of this class [16].

Smoothness functionals that can be expressed as weighted norms over the Fourier transform domain are unable to characterize spatially varying smoothness information. As a result, a regularization approximation associated with the class of smoothness prior defined at (4.1) consists of a smooth function fitted to the data set, with energy localized around a spatial frequency of zero. Moreover, the optimal approximant in this case comprises (in addition to a term that differentiates between functions in the null space of K) a linear combination of basis functions g whose centres lie on

the sample points $(x_i)_{i=1}^N$, and therefore is very redundant in general.

On the other hand, the rate of decay across scale of the magnitude of a function's wavelet transform characterizes the local regularity of the function to a degree dependent on the regularity of the analysing wavelet [25, 22, 21]. This observation suggests the use of a weighted norm over the wavelet transform domain as a smoothness functional, with a weighting factor constructed so that K is minimized by a function having energy contained in regions of the space-frequency domain in which the energy of the data-set is localized. Moreover, if the approximation is a truncated expansion over a wavelet family, the localization of $\hat{f}$ to regions of the space-frequency domain is governed by the characteristics of the basis functions incorporated in the expansion. A smoothness functional of this kind can therefore be interpreted as a measure of the degree to which the basis function set is matched to the wavelet transform of the approximated function.

Let ϕ be a real-valued wavelet satisfying an admissibility condition $C_\phi < \infty$ [8], and denote by Wf the wavelet transform with respect to ϕ, defined over all real-valued functions $f \in L^2(\mathcal{R})$. Consider the functional:

$$K[\hat{f}] = C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty db \, \frac{[W\hat{f}(a,b)]^2}{\sum_{\alpha,\beta \in \mathcal{A}} (2 + < \phi^{\alpha,\beta}, \phi^{a,b} >)}, \tag{4.2}$$

where the set $\mathcal{A}$ contains a finite and non-zero number of indices. By defining a second wavelet ϕ_a, containing the positive frequency components of ϕ alone ($\mathcal{F}\phi_a(\xi) = \mathcal{F}\phi(\xi)$ for $\xi > 0$, and $\mathcal{F}\phi(\xi) = 0$ otherwise), each term in the sum in the denominator of (4.2) can be rewritten as $\|\phi_a^{\alpha,\beta} + \phi_a^{a,b}\|^2$ (ϕ_a is normalized so that $\|\phi_a^{a,b}\| = 1$). The denominator therefore has a greatest lower bound $1/U_{\mathcal{A}}$, strictly positive for non-empty $\mathcal{A}$. Since in addition, the number of index pairs in $\mathcal{A}$ is finite, there exists a least upper bound $1/L_A$ on the denominator due to the bound $2 + < \phi^{\alpha,\beta}, \phi^{a,b} > \leq 4$ that holds on each term in the sum. Making use of the isometric property of the continuous wavelet transform, it follows that

$$L_{\mathcal{A}} \|f\|^2 \leq K[f] \leq U_{\mathcal{A}} \|f\|^2. \tag{4.3}$$

The smoothness measure $K[f]$ is therefore well defined for all $f \in L^2(\mathcal{R})$, and $K[f] = 0$ if and only if $f(\mathbf{x}) = 0$ identically.

Let the set J contain the indices of coefficients that exceed a threshold λ in an expansion of the approximated function over the discretely parameterized family $(\phi_j)_{j\in\mathcal{Z}}$. If $\mathcal{A}$ contains the points in the continuous ab-domain corresponding to the centres of localization of wavelets in the set $(\phi_j)_{j\in J}$, then the wavelet transform of the minimizing argument of K is concentrated within regions at which the continuous wavelet transform of the data set exceeds λ. This can be seen by considering the function f that minimizes K subject to the constraint $\|f\|^2 = 1$. Using Lagrange's method, f is the minimizing argument of a functional

$$H[f] = K[f] + \lambda_1(\|f\|^2 - 1),$$

for some $\lambda_1 \in \mathcal{R}$.

Let $(c_j)_{j\in\mathcal{Z}}$ be the coefficients in the expansion of f over an orthonormal wavelet basis $(e_j)_{j\in\mathcal{Z}}$, then, with $K[e_j, e_k]$ defined by

$$K[e_j, e_k] = C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty db \, \frac{We_j(a,b)We_k(a,b)}{\sum_{\alpha,\beta\in\mathcal{A}}(2 + < \phi^{\alpha,\beta}, \phi^{a,b} >)},$$

the bounds $L_{\mathcal{A}} | < e_j, e_k > | \leq |K[e_j, e_k]| \leq U_{\mathcal{A}} | < e_j, e_k > |$ give $K[e_j, e_k] = 0$ for all $j \neq k$, so that H can be written as the following function of $(c_j)_{j\in\mathcal{Z}}$

$$H[(c_j)_{j\in\mathcal{Z}}] = \sum_{j\in\mathcal{Z}} c_j^2 K[e_j] + \lambda_1 \left[\sum_{j\in\mathcal{Z}} c_j^2 - 1 \right].$$

Owing to the continuity of H in each of the parameters c_j, $\partial H/\partial c_j = 0 \; \forall j \in \mathcal{Z}$ must hold at the extremum points of H. These conditions imply that $c_j(K[e_j] + \lambda_1) = 0 \; \forall j \in \mathcal{Z}$, so that $\lambda_1 = -K[e_{j_0}]$ for some j_0 and $c_j = 0 \; \forall j \neq j_0 \in \mathcal{Z}$. It follows that all local extrema of $K[f]$ on the space of functions with unit L^2-norm are given by

$f = e_j$ for all $j \in \mathcal{Z}$, so that the extremum values of K are given by

$$K[e_j] = C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty db \, \frac{[We_j(a,b)]^2}{\sum_{\alpha,\beta \in \mathcal{A}} (2 + < \phi^{\alpha,\beta}, \phi^{a,b} >)}.$$

Since each of the terms $< \phi^{\alpha,\beta}, \phi^{a,b} >$ in the denominator of K is a maximum with respect to (a,b) at (α,β) and decays to zero asymptotically with distance from this point, the localization of e_j about the point in ab-space specified by j suggests that the global minimum of $K[f]$ is obtained with $f = e_j$ for some $j \in J$. As a result, K represents the information that the energy of the data set $(x_i, f(x_i))_{i=1}^N$ is concentrated at the points in the space-frequency domain specified by the set J.

In appendix 4.3, it is shown that a regularization cost functional employing the smoothness measure (4.2) is minimized by a function having the form

$$\hat{f}(x) = \sum_{\alpha,\beta \in \mathcal{A}} c_{\alpha,\beta} \phi^{\alpha,\beta}(x), \ \forall x \neq x_i,$$

where ϕ is a member of the class of wavelets defined by (4.9), and $(c_{\alpha,\beta})_{\alpha,\beta \in \mathcal{A}}$ are undetermined coefficients. Since $\mathcal{A}$ contains the points in the continuous ab-domain that correspond to the indices contained in J, this function can be written as an expansion over the discretely parameterized wavelets indexed by J:

$$\hat{f}(x) = \sum_{j \in J} c_j \phi_j(x), \ \forall x \neq x_i. \tag{4.4}$$

The exclusion of the sample points $(x_i)_{i=1}^N$ from the domain on which $\hat{f}$ is defined is a consequence of the restriction of the domain of the wavelet transform to the space $L^2(\mathcal{R})$. Provided the scaling parameters of the wavelets indexed by J are finite, the sum $\sum_{j \in J} c_j \phi_j(x)$ on the right side of equation (4.4) is uniformly continuous. For practical purposes therefore, the expression for $\hat{f}$ can be extended to all values of $x \in \mathcal{R}$.

The optimal approximant with respect to the smoothness measure defined at (4.2) is therefore a truncated expansion over a set of wavelets localized at points in space and frequency determined by the smoothness functional alone. In contrast, a smooth-

ness prior of the class given by (4.1) specifies an approximant consisting of a linear combination of basis functions g that are centred on the sample points $(x_i)_{i=1}^N$ [16]. Since the number and locations of basis functions required for a given approximation accuracy is dependent on the regularity of the approximated function rather than the distribution of sample points, the wavelet basis function approximation derived above is likely to be more efficient than regularization approximations obtained using the class (4.1) of smoothness priors based on the Fourier transform.

The estimated coefficients in a network approximation of the form of (4.4) to a function $f_A(x)$ can be considered as estimates of the inner products $< f_A, \tilde{\phi}_j >$ of the approximated function with the elements of the dual frame. Moreover, the inner products of f_A with the reconstructing wavelets $< f_A, \phi_j >$ approximate reasonably well the inner products of f_A with the members of the dual frame for $j \in J$, where J is a finite index set [8]. As a result, the coefficients in the network expansion serve as estimates of the values of the continuous wavelet transform of f_A at points in the continuous ab-domain corresponding to the discretely parameterized scale and translation parameters of their associated basis functions. The regularization approximation derived above therefore justifies retaining in a truncated expansion only the coefficients exceeding a threshold λ in magnitude. Using this strategy, a spatially adaptive approximation $\hat{f}$ is effectively obtained by adapting K according to the smoothness information that becomes available during the process of constructing the approximation.

Although the preceding discussion has been restricted to approximation on single-dimensional spaces, the regularization approach described above can be extended to the multi-dimensional case, for example by taking tensor products of both continuous and discretely parameterized wavelet families.

4.3 Appendix: Minimization Of The Regularization Cost Functional

Using the smoothness functional defined at (4.2), the regularization cost functional can be expressed in terms of the continuous wavelet transform of the approximant $\hat{f}$ as follows

$$P[W\hat{f}] = \sum_{i=1}^{N}\left[C_\phi^{-1}\int_0^\infty \frac{da}{a^2}\int_{-\infty}^{\infty} db\, W\hat{f}(a,b)\phi^{a,b}(x_i) - f(x_i)\right]^2 + \lambda_0 C_\phi^{-1}\int_0^\infty \frac{da}{a^2}\int_{-\infty}^{\infty} db\, \frac{[Wf(a,b)]^2}{\sum_{\alpha,\beta\in\mathcal{A}}(2+ <\phi^{\alpha,\beta},\phi^{a,b}>)}.$$

Define a function $F(a,b) = W\hat{f}(a,b) + \epsilon\eta(a,b)$, where $\eta(a,b)$ is an arbitrary real-valued function in $\mathcal{H}$ the image space of W, and ϵ is a real scalar. The effect of perturbing $W\hat{f}$ on the value of the cost functional can be determined by evaluating $P[F]$:

$$P[F] = \sum_{i=1}^{N}\left[C_\phi^{-1}\int_0^\infty \frac{da}{a^2}\int_{-\infty}^{\infty} db\, [W\hat{f}(a,b)+\epsilon\eta(a,b)]\phi^{a,b}(x_i) - f(x_i)\right]^2 + \lambda_0 C_\phi^{-1}\int_0^\infty \frac{da}{a^2}\int_{-\infty}^{\infty} db\, \frac{[W\hat{f}(a,b)+\epsilon\eta(a,b)]^2}{\sum_{\alpha,\beta\in\mathcal{A}}(2+ <\phi^{\alpha,\beta},\phi^{a,b}>)}.$$

Since the cost functional is continuous in F, it is necessarily stationary with respect to variations $\epsilon\eta$ in functions $W\hat{f}$ that furnish extrema. A necessary condition that $W\hat{f}$ minimizes P is therefore given by $\partial P[F]/\partial\epsilon|_{\epsilon=0} = 0$. Differentiating with respect to ϵ under the integral sign, and imposing the condition $\partial P[F]/\partial\epsilon|_{\epsilon=0} = 0$, we have

$$C_\phi^{-1}\int_0^\infty \frac{da}{a^2}\int_{-\infty}^{\infty} db\left[\sum_{i=1}^{N}[\hat{f}(x_i)-f(x_i)]\phi^{a,b}(x_i) + \lambda_0\frac{W\hat{f}(a,b)}{\sum_{\alpha,\beta\in\mathcal{A}}(2+ <\phi^{\alpha,\beta},\phi^{a,b}>)}\right]\eta(a,b) = 0.$$

This equation must hold for all η in $\mathcal{H}$, implying that:

$$W\hat{f}(a,b) = -\lambda_0^{-1}\sum_{\alpha,\beta\in\mathcal{A}}\sum_{i=1}^{N}[\hat{f}(x_i)-f(x_i)](2+ <\phi^{\alpha,\beta},\phi^{a,b}>)\phi^{a,b}(x_i),\ \text{a.e.} \qquad (4.5)$$

The complex wavelet ϕ_a defined in section 4.2 has an associated wavelet transform $W_a f$, related to the wavelet transform with respect to ϕ by $Wf(a,b) = W_a f(a,b) + \overline{W_a f(a,b)}$ (where $\overline{W_a f}$ denotes the complex conjugate of $W_a f$). The expression (4.5) for the wavelet transform of the function that minimizes P is therefore equivalent to a condition on the real part of $W_a \hat{f}$:

$$\begin{aligned} Re(W_a \hat{f}(a,b)) &= -\lambda_0^{-1} \sum_{\alpha,\beta \in \mathcal{A}} \sum_{i=1}^{N} [\hat{f}(x_i) - f(x_i)] \\ &\quad .(2 + < \phi_a^{\alpha,\beta}, \phi_a^{a,b} > + \overline{< \phi_a^{\alpha,\beta}, \phi_a^{a,b} >}) Re(\phi_a^{a,b}(x_i)). \end{aligned}$$

A sufficient condition on $W_a \hat{f}$ is given by

$$W_a \hat{f}(a,b) = -\lambda_0^{-1} \sum_{\alpha,\beta \in \mathcal{A}} \sum_{i=1}^{N} [\hat{f}(x_i) - f(x_i)](2 + < \phi_a^{\alpha,\beta}, \phi_a^{a,b} > + \overline{< \phi_a^{\alpha,\beta}, \phi_a^{a,b} >}) \phi_a^{a,b}(x_i). \tag{4.6}$$

Since the null space of the smoothness functional $K[\hat{f}]$ contains only $\hat{f} = 0$, the solution to the minimization problem is unique [20]. It follows that the function $\hat{f}$ that satisfies a sufficient condition on the minimizing argument of P is the unique optimal approximant given the regularization cost functional.

In order to recover $\hat{f}$, the wavelet transform in equation (4.6) must be inverted. Inversion of $W_a \hat{f}$ with respect to ϕ_a yields a function $\hat{f}_a$ containing only the positive frequency components of $\hat{f}$, so that $\mathcal{F}\hat{f}_a(\xi) = \mathcal{F}\hat{f}(\xi)$ for $\xi > 0$ and $\mathcal{F}\hat{f}_a(\xi) = 0$ for $\xi < 0$. From (4.6), it follows that

$$\begin{aligned} \hat{f}_a(x) = -\lambda_0^{-1} \sum_{\alpha,\beta \in \mathcal{A}} \sum_{i=1}^{N} [\hat{f}(x_i) - f(x_i)] \qquad (4.7) \\ .C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^{\infty} db\, \phi_a^{a,b}(x_i) \phi_a^{a,b}(x) (2 + < \phi_a^{\alpha,\beta}, \phi_a^{a,b} > + \overline{< \phi_a^{\alpha,\beta}, \phi_a^{a,b} >}). \end{aligned}$$

The first term in the above integral can be evaluated as

$$\begin{aligned} C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^{\infty} db\, \phi_a^{a,b}(x_i) \phi_a^{a,b}(x) &= C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^{\infty} d\xi\, a e^{i\xi(x-x_i)} \mathcal{F}\phi_a(-a\xi) \mathcal{F}\phi_a(a\xi) \\ &= 0, \end{aligned}$$

where $\mathcal{F}\phi_a(\xi) = 0$ for $\xi < 0$ has been used. Let the Fourier transform of $\phi^{a,b}(x_i)\phi^{a,b}(x)$ with respect to the variable b be $G(\xi)$. Then, since support$G \subset (-\infty, 0]$, and $\mathcal{F}\phi_a(\xi)\mathcal{F}\phi^{\alpha,\beta}(\xi) = 0$ for $\xi < 0$, the third term in the integral of equation (4.7) evaluates to

$$\begin{aligned} C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty db\, \phi_a^{a,b}(x_i)\phi_a^{a,b}(x)\overline{< \phi_a^{\alpha,\beta}, \phi_a^{a,b} >} \\ &= \frac{(2\pi)^{1/2}}{C_\phi} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty d\xi\, G(\xi)\overline{[a^{1/2} \exp^{i\xi b}\, \mathcal{F}\phi^{\alpha,\beta}(\xi)\mathcal{F}\phi(a\xi)]} \\ &= 0. \end{aligned}$$

Equation (4.7) therefore reduces to

$$\hat{f}_a(x) = -\lambda_0^{-1} \sum_{\alpha,\beta \in \mathcal{A}} \sum_{i=1}^{N} [\hat{f}(x_i) - f(x_i)] C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty db\, < \phi_a^{\alpha,\beta}, \phi_a^{a,b} > \phi_a^{a,b}(x_i)\phi_a^{a,b}(x). \tag{4.8}$$

Let $\mathcal{H}$ be the image space of the wavelet transform W_a. The integral in equation (4.8) can be rewritten as the inner product in $\mathcal{H}$ of $\overline{\phi_a^{a,b}(x)\phi_a^{a,b}(x_i)}$ with the reproducing kernel $K_{\alpha,\beta}(a,b)$ of $\mathcal{H}$. From the reproducing kernel property [1], it follows that

$$C_\phi^{-1} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty db\, < \phi_a^{\alpha,\beta}, \phi_a^{a,b} > \phi_a^{a,b}(x_i)\phi_a^{a,b}(x) = \overline{\Phi(\alpha,\beta)},$$

where $\Phi(a,b)$ is the orthogonal projection of $\overline{\phi_a^{a,b}(x)\phi_a^{a,b}(x_i)}$ onto the space $\mathcal{H}$ in the case that $\overline{\phi_a^{a,b}(x)\phi_a^{a,b}(x_i)}$ is not a member of $\mathcal{H}$, and $\Phi(a,b)$ is simply $\overline{\phi_a^{a,b}(x)\phi_a^{a,b}(x_i)}$ otherwise.

An arbitrary $L^2(\mathcal{R}^2)$ function cannot in general be assumed to belong to the image space of W_a, since $\mathcal{H}$ is embedded as a subspace in, but not necessarily equal to $L^2(\mathcal{R}^2)$. However, if the analysis is restricted to a particular class of wavelet having Fourier transform

$$\mathcal{F}\phi_a(\xi) = \begin{cases} C_n \xi^n \exp^{-\xi} & \xi > 0 \\ 0 & \xi < 0 \end{cases}, \quad C_n = \frac{2^{n+1/2}}{(J(2n+1))^{1/2}}, \tag{4.9}$$

(where the factor C_n is included so that $\|\phi_a\| = 1$) then $a^{-n-1/2}F(a,b)$ is analytic in

the variable $b+ia, a > 0$ for all $F \in \mathcal{H}$. In this case, the space $\{a^{-n-1/2}F(a,b); F \in \mathcal{H}\}$ is identical to the space of functions analytic and square-integrable with respect to the measure $a^{2n-1}da\,db$ on the upper half of the complex plane [8]. Since $a^{-n-1/2}\overline{\phi_a^{a,b}(x)}$ is analytic in the variable $b + ia$, $a > 0$, the product $a^{-2n-1}\overline{\phi_a^{a,b}(x)\phi_a^{a,b}(x_i)}$ is necessarily also analytic. Furthermore, it can be verified that $a^{-2n-1}\overline{\phi_a^{a,b}(x)\phi_a^{a,b}(x_i)}$ is in $L^2(\mathcal{R}^2; a^{2n-1}da;\ db)$ for all $x \neq x_i$, allowing the conclusion that $\overline{\phi_a^{a,b}(x)\phi_a^{a,b}(x_i)}$ is a member of $\mathcal{H}$ provided $x \neq x_i$. For this class of wavelet, $\hat{f}_a(x)$ can therefore be obtained for $x \neq x_i$ from equation (4.8) as

$$\hat{f}_a(x) = -\lambda_0^{-1} \sum_{\alpha,\beta \in \mathcal{A}} \sum_{i=1}^{N} [\hat{f}(x_i) - f(x_i)]\phi_a^{\alpha,\beta}(x_i)\phi_a^{\alpha,\beta}(x)$$

The minimizing argument of the cost functional P then follows from $\hat{f} = 2Re(\hat{f}_a)$:

$$\hat{f}(x) = \sum_{\alpha,\beta \in \mathcal{A}} c_{\alpha,\beta}\phi^{\alpha,\beta}(x)$$

for all $x \neq x_i$ and some set of coefficients $(c_{\alpha,\beta})_{\alpha,\beta \in \mathcal{A}}$.

Chapter 5

Trajectory Estimation For A Robotic Catching System

This chapter describes an approach to predicting on-line the path of an object moving freely through air using measurements of the position and velocity of its centre of mass. The system was developed with the objective of accurately estimating the trajectories of a broad class of objects with a minimum of prior knowledge of object geometries and aerodynamic properties. As a result, fixed, predetermined models of the aerodynamic forces experienced by the object are unsuited to this application. Instead, the dynamics of the object's motion are identified empirically over the course of a single trajectory using the identification network developed in this thesis.

The network is trained on measurements of the object's position and velocity obtained from the images of a pair of CCD cameras. Estimates of the object's future positions, velocities and accelerations generated using the network model are used to specify the trajectory of the end point of a robot arm that is to be coordinated to catch the moving object. The high degree of accuracy in predictions of the object's state that this task requires motivates the use of a pair of moving cameras with narrow fields of view, rather than fixed cameras with wide fields of view. With the cameras mounted on robotic actuators, their angular orientations and velocities can be accurately measured, so that the limiting effect of finite image resolution on the system's measurement accuracy is reduced.

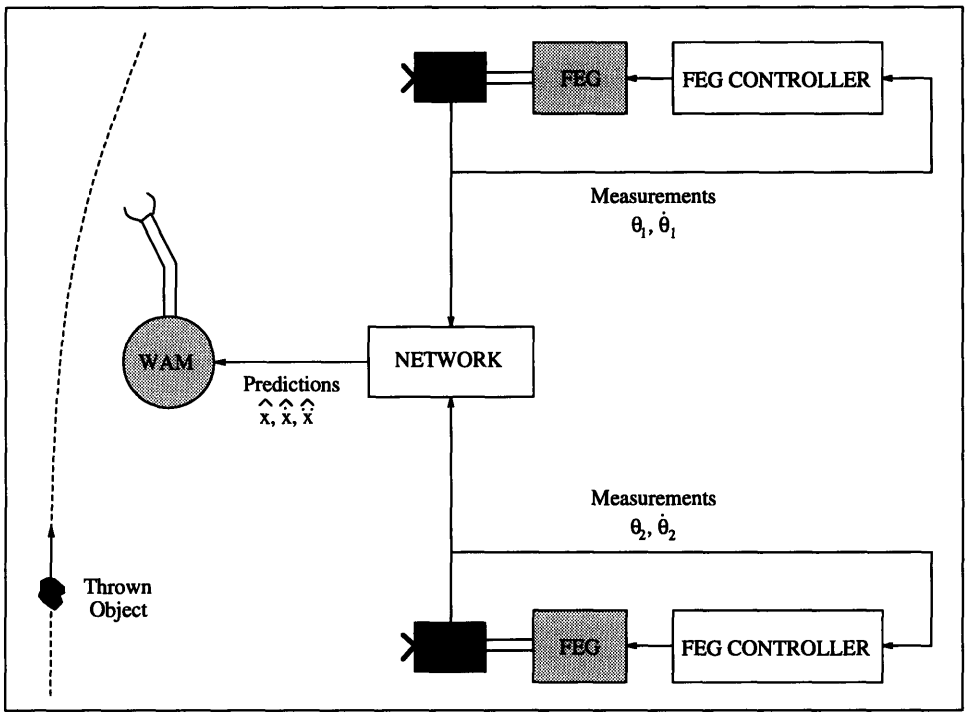

Figure 5-1: A diagram of the catching system showing the 6 degree-of-freedom cable-driven Whole Arm Manipulator (WAM) that is to catch the thrown object, and 2 degree-of-freedom, cable-driven Fast Eye Gimbal (FEG) camera actuators.

The images of the two cameras are processed independently of each other. Each camera supplies images at a rate of 60Hz. In order to process images at this rate, a relatively simple object detection algorithm is implemented. The image processor of each camera recognizes an object on the basis of its colour, and generates position measurements $(\boldsymbol{\theta}_i(t), \dot{\boldsymbol{\theta}}_i(t),\ i = 1, 2)$ by estimating its centre of mass as the centre of area of the region of that colour in the image. The measurements of the object's velocity are obtained through numerical differentiation of position measurements. No measurements of the object's angular orientation are made.

The actuators on which the cameras are mounted are independently controlled. The desired trajectory to be followed by each actuator is generated from its associated image processor's estimates $\boldsymbol{\theta}_i(t), \dot{\boldsymbol{\theta}}_i(t)$ of the object's current position and velocity in the space of the individual actuator's joint rotations. To compensate for a delay of approximately 50ms between the acquisition of an image and the availability of new state information from the image processor, state measurements are extrapolated on

the assumption of constant object velocity over the delay interval. Between vision samples, the object's current state in the joint space of each camera actuator is also estimated by extrapolating position measurements assuming the object's velocity to remain constant.

Measurements from the two image processors are used to calculate the object's position and velocity $(\mathbf{x}(t), \dot{\mathbf{x}}(t))$ in a fixed Cartesian reference frame. This information is used to update the network model each time new measurements of the object's state become available. Over the remainder of the interval between vision samples the network continually generates predictions of the object's future positions, velocities and accelerations $\hat{\mathbf{x}}(t+T), \hat{\dot{\mathbf{x}}}(t+T), \hat{\ddot{\mathbf{x}}}(t+T)$ on the basis of the state estimates from which the desired trajectories for the camera actuators are generated.

The following section describes a simple model of the structure of the object's dynamics that allows rapid network training and reconstruction. Although the simplifying assumptions made in the derivation of this model reduce the generality of the network's approximation, the errors in state predictions that result are small relative to those caused by measurement noise, provided the modeling assumptions hold approximately. This is demonstrated by the experimental results given in section 5.2.

5.1 Network Design

In order to keep to a minimum the time required to update and reconstruct the network's approximation, the dimensionality of its input space must be minimized. This section develops a model of the structure of the dynamics of a free-falling object that allows its state to be predicted over the course of a single trajectory using single-dimensional networks. Coefficient adaptation laws that allow stable identification of the object's accelerations, and asymptotic bounds on the convergence of errors in the network-based estimates are given. The generality of the class of objects whose dynamics can be accurately represented by this model is discussed, along with restrictions on the method of generating state predictions.

At time t, the object's orientation relative to its direction of motion is given by

the two-dimensional vector $\boldsymbol{\gamma}(t)$, and its centre of mass is $\mathbf{x}(t)$ in a fixed Cartesian reference frame. Assuming it to be moving through still air of uniform density and viscosity, the aerodynamic forces experienced by the object are dependent on the time-varying quantities $v = \|\dot{\mathbf{x}}(t)\|$, $\boldsymbol{\gamma}(t)$ and $\dot{\boldsymbol{\gamma}}(t)$ alone. The object's acceleration may therefore be written

$$\ddot{\mathbf{x}}(t) = \mathbf{F}^t(v(t), \boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t)) + \mathbf{F}^n(v(t), \boldsymbol{\gamma}(t), \dot{\boldsymbol{\gamma}}(t)) + \mathbf{g},$$

where the aerodynamic accelerations in directions tangential and normal to the object's trajectory are $\mathbf{F}^t$ and $\mathbf{F}^n$ respectively, and the acceleration due to gravity is $\mathbf{g}$.

Provided the attitude $\boldsymbol{\gamma}$, or, if the object is rotating about an axis of rotational symmetry, its time-derivative $\dot{\boldsymbol{\gamma}}$ is approximately constant over the duration of a trajectory, the aerodynamic forces can be idealized as $\mathbf{F}^t = \mathbf{F}^t(v)$ and $\mathbf{F}^n = \mathbf{F}^n(v)$. In this case the direction of $\mathbf{F}^n$, characterized by the angle γ between the line of intersection of the plane normal to the direction of motion and the xy-plane (see figure 5-2), is constant, and velocity measurements specify v and the direction of $\mathbf{F}^t$. Measurements of $\dot{\mathbf{x}}$ therefore allow identification of the the dynamics of the object's motion in a trajectory-based coordinate system by three single-dimensional networks $\hat{f}^t(v)$, $\hat{f}^{nc}(v)$, and $\hat{f}^{ns}(v)$, providing approximations to $\|\mathbf{F}^t(v)\|$, $\|\mathbf{F}^n(v)\| \cos\gamma$, and $\|\mathbf{F}^n(v)\| \sin\gamma$.

The object's acceleration $\mathbf{f} = [\|\mathbf{F}^t\| \;\; \|\mathbf{F}^n\| \cos\gamma \;\; \|\mathbf{F}^n\| \sin\gamma]^T$ due to aerodynamic forces in the trajectory-based coordinate system is related to its acceleration $\mathbf{F}$ in the Cartesian reference frame by the transformation $\mathbf{F} = \mathbf{R}\mathbf{f} + \mathbf{g}$, where the rotation matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \frac{1}{v v_{xy}} \begin{bmatrix} \dot{x} v_{xy} & \dot{y} v & -\dot{x}\dot{z} \\ \dot{y} v_{xy} & -\dot{x} v & -\dot{y}\dot{z} \\ \dot{z} v_{xy} & 0 & v_{xy}^2 \end{bmatrix},$$

and $v_{xy}^2 = (\dot{x})^2 + (\dot{y})^2$. An estimate $\hat{\mathbf{F}} = \mathbf{R}\hat{\mathbf{f}} + \mathbf{g}$ of the object's Cartesian acceleration vector can therefore be generated from the vector $\hat{\mathbf{f}} = [\hat{f}^t \;\; \hat{f}^{nc} \;\; \hat{f}^{ns}]^T$ of network outputs.

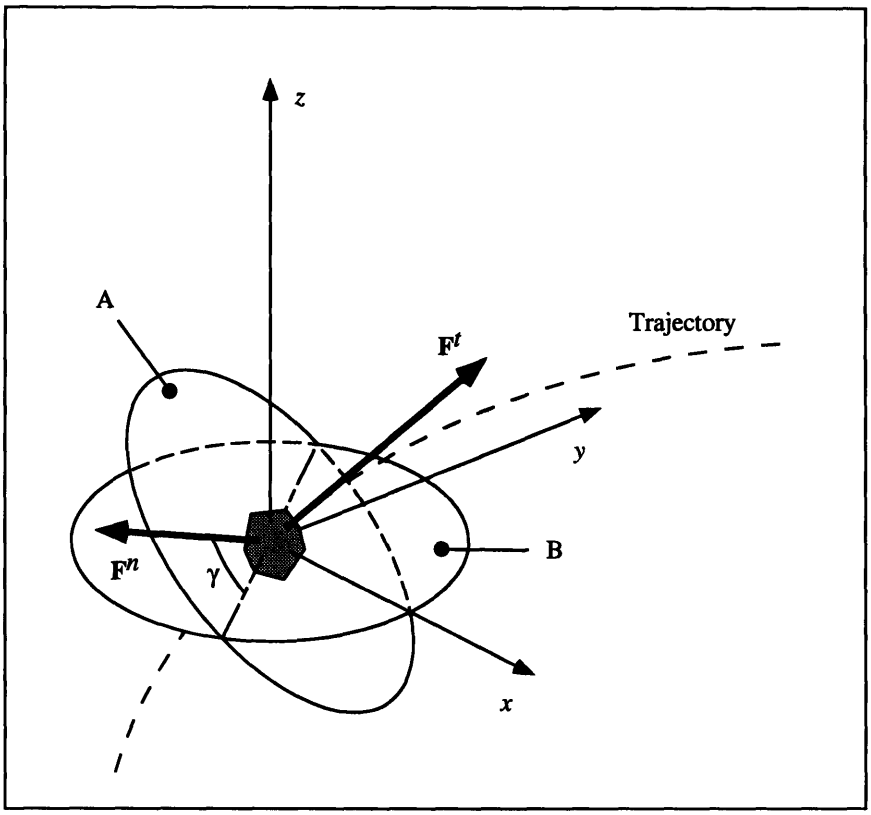

Figure 5-2: Accelerations $\mathbf{F}^t$ and $\mathbf{F}^n$ due to the aerodynamic forces experienced by the object. The plane normal to the trajectory is represented by A, and B represents the xy-plane. The angle γ is in the plane of A.

Using the identifier described in section 2.3, the networks are trained on an error measure $\mathbf{e}(t) = \hat{\mathbf{v}}(t) - \dot{\mathbf{x}}(t)$, with $\hat{\mathbf{v}}$ evolving according to

$$\dot{\hat{\mathbf{v}}} = -k_D\mathbf{e} + \hat{\mathbf{F}}, \tag{5.1}$$

where $k_D > 0$ is a scalar constant. In terms of the error in the reconstruction $\hat{\mathbf{f}} - \mathbf{f}$, the dynamics of the error measure are given by

$$\mathbf{R}^T\dot{\mathbf{e}} = -k_D\mathbf{R}^T\mathbf{e} + \hat{\mathbf{f}} - \mathbf{f}. \tag{5.2}$$

In the trajectory based coordinate system, the error measure can be written $\mathbf{R}^T\mathbf{e} =$

$[e^t\ e^{nc}\ e^{ns}]^T$, and a Lyapunov-like function $V_{\mathbf{e}}$ can be defined

$$V_{\mathbf{e}} = V_{e^t} + V_{e^{nc}} + V_{e^{ns}},$$

where V_{e^t}, $V_{e^{nc}}$, and $V_{e^{ns}}$ are given by equation 2.16 with e replaced by e^t, e^{nc}, and e^{ns} respectively. Since $\mathbf{e}^T\mathbf{e} = (e^t)^2 + (e^{nc})^2 + (e^{ns})^2$, the analysis of section 2.4 can be applied to $V_{\mathbf{e}}$ and the error dynamics (5.2) to show the global stability of the identifier, using the structural adaptation laws described in section 2.1 and the output weight adaptation laws of equation 2.15 with e replaced by e^t, e^{nc}, and e^{ns} for the networks generating f^t, f^{nc} and f^{ns} respectively. The argument of section 2.5 then gives the following upper bound on the asymptotic convergence of the error measure

$$\limsup_{T\to\infty} \frac{1}{T}\int_t^{t+T} \mathbf{e}^T(\tau)\mathbf{e}(\tau)d\tau \leq \frac{1}{k_D^2} \sup_{\tau\in[t,t+T]} \boldsymbol{\epsilon}^T(\mathbf{x}(\tau),\tau)\boldsymbol{\epsilon}(\mathbf{x}(\tau),\tau),$$

and the vector $\boldsymbol{\epsilon}(\mathbf{x}(t),t)$ of minimal approximation errors in the above expression is defined at equation 2.10, with $\hat{f}$ and f replaced by $\hat{\mathbf{f}}$ and $\mathbf{f}$.

For rotationally symmetric objects, the assumption that $\dot{\gamma}$ is approximately constant requires that the object's angular velocity vary slowly. This condition is generally satisfied if the angular momentum is sufficiently high that the associated aerodynamic force $\mathbf{F}^n$ is comparable to gravity. The assumption that an irregularly shaped object's attitude remains constant holds approximately if the object is aerodynamically stable.

The simplifying assumptions that allow the object's dynamics to be approximated by functions of a single variable impose the additional restriction that delays cannot be directly incorporated in the network model. This is a consequence of the change in the direction of motion of the object over the delay interval, which has the effect that an error measure generated using the measured velocity $\dot{\mathbf{x}}(t_1)$ and the reconstruction $\hat{\mathbf{f}}(v(t_0))$ in a coordinate system based on the trajectory at time t_0, cannot be used to train the network. As a result, accurate estimates of the object's acceleration at time t_1 cannot be obtained from network reconstructions based on $v(t_0)$, if $t_1 - t_0$ is comparable to the time-scale of changes in the direction of the trajectory. The pre-

dictions $\hat{\mathbf{x}}(t_1)$, $\hat{\dot{\mathbf{x}}}(t_1)$, and $\hat{\ddot{\mathbf{x}}}(t_1)$ made at time t_0 are therefore generated by integrating the network's acceleration estimates $\hat{\ddot{\mathbf{x}}}(t)$ over the interval $t_0 \leq t \leq t_1$.

5.2 Experimental Results

This section presents experimental results illustrating the performance of the network-based identification scheme described in section 5.1. The experimental implementation of the network-based identifier is described, and network predictions are contrasted with the trajectory predictions of simple non-adaptive models of the thrown object's dynamics. Several objects with differing aerodynamic properties were used; a smooth heavy ball, a light sponge ball, and a paper aeroplane.

In each experiment, the three networks were implemented in software on a pair of TMS320C40 digital signal processors. The larger of the three networks (that generating $\hat{f}^t$ in the light and heavy ball identification experiments, and that generating $\hat{f}^{ns}$ in the paper aeroplane identification experiment) was run in parallel with the remaining two networks. The two DSP chips were synchronized so that all three networks were updated on a single set of measurements $\mathbf{x}(t), \dot{\mathbf{x}}(t), \ddot{\mathbf{x}}(t)$, and the network model $\hat{\mathbf{F}}(v)$ reconstructed on the basis of a single velocity $v(t)$.

Table 5.1 summarizes the parameters of the identifiers used in the three experiments. Identical system parameters were used in the light and heavy ball identification experiments, and in individual experiments, the network parameters λ, μ, $|f|_{\max}$, k_a, and σ_0 were equal for the three networks. From the error measure dynamics (5.2), it can be seen that the value of k_D determines the bandwidth of the error measure $\mathbf{e}(t)$, which can be considered as a filtered version of the reconstruction error $\hat{\mathbf{F}}(t) - \mathbf{F}(t)$. A cut-off frequency $k_D = 10$Hz was used in each of the experiments in order to smooth the component of measurement noise at 30Hz while allowing a 90% rise-time in the error measure of 0.25s.

An indication of the order of magnitude of the adaptive gain k_a to be used in each experiment was obtained from the maximum aerodynamic force $|f|_{\max}$, and estimates of the magnitude of noise $\|\mathbf{e}_v\|$ in the measurements of object velocity $\dot{\mathbf{x}}(t)$. The rates

	λ (ms^{-2})	μ (ms^{-2})	$\|f\|_{max}$ (ms^{-2})	$\|A_T\|$ (ms^{-1})	p_λ
A	0.05	0.001	2.0	5.0	9.77
B	0.1	0.001	10.0	5.0	12.4

	k_D (s^{-1})	k_a (s^{-1})	σ_0 (s^{-1})
A	10.0	2.0	2.0
B	10.0	0.5	2.0

Table 5.1: Experimental network parameters. A: Parameters used in the identification of the dynamics of the light and heavy balls. B: Parameters used in the identification of the paper aeroplane's dynamics.

of change of coefficient estimates $\dot{\hat{c}}_j$ are of order $k_D|f|_{max}$, and the component of $\dot{\hat{c}}_j$ due to the measurement noise $\|\mathbf{e}_v\|$ is of order $k_a k_D \|\mathbf{e}\|_v$. The condition that the relative component of estimates' derivatives $\dot{\hat{c}}_j$ due to measurement noise is small gave a bound on the order of magnitude of the adaptive gain: $k_a \ll |f|_{max}/\|\mathbf{e}\|_v$.

Figures 5-3, 5-4 and 5-5 compare $\|\dot{\mathbf{x}}(t)\|$ and $\|\hat{\mathbf{v}}(t)\|$; the magnitudes in each experiment of the measured object velocity and the velocity estimate used to generate the identifier error measure respectively. The error measures of the light and heavy ball identifiers converged to their steady-state values within 0.25s. The error measure converged more slowly in the paper aeroplane identification experiment since a reduced adaptive gain was used in order to compensate for greater errors in velocity measurements. The higher levels of measurement noise in the paper aeroplane identification experiment resulted from an increase in the sensitivity of estimates of the object's centre of mass to object orientation and lighting conditions in this case.

Figures 5-6 to 5-8 show for the three experiments the steady-state values of network coefficients and the locations of the centres of their associated basis functions on the lattice of velocity-frequency localization points. The apparent elongation of basis functions centred at fine scales in these plots is due to the close spacing of nodes in the direction of varying v. The networks approximating f^t in the ball identification experiments contained greater numbers of nodes, and their coefficients were generally larger than those in the networks approximating f^{nc} and f^{ns}. This indicates that the drag force identified by the network model was more significant than forces acting normal to the object's trajectory in these cases. In addition, this characteristic is

accentuated in the model of the light ball's aerodynamic forces, agreeing with the expectation that drag forces are more significant in the light ball's dynamics.

The networks modelling the paper aeroplane's dynamics contained more nodes and coefficients of greater magnitudes than the corresponding networks identifying the dynamics of either of the balls, indicating that the aerodynamic forces experienced in this case were more significant in relation to gravitational forces. Moreover, the network approximating f^{ns} contained the coefficients greatest in magnitude and the largest number of nodes of the three networks. Since the plane's trajectory was near-horizontal in this experiment, this shows that the lift force identified by the network was more significant than drag forces and forces normal to the plane's trajectory in directions other than vertical.

The values of $|f|_{\max}$ used in each experiment were based on rough estimates of the maximum aerodynamic force experienced by the object. The value of $0.01|f|_{\max}$ was taken as an indication of the order of magnitude of λ required in order to obtain identifier models consisting of approximately 50 steady-state nodes distributed between the three networks. The larger value of λ used in the paper aeroplane identification experiment reflects the larger aerodynamic forces experienced by this object. The candidate basis function set of each network consisted of a family of mexican hat wavelets constituting an approximately tight frame ($m = 6.55$, $M = 7.09$). Since basis functions were normalized so that $\|\phi\| = 1$, the value of $|\tilde{\phi}|_{\max}$ in this case was approximately 0.147 [8]. As a result, p_λ was approximately equal to 10 in each of the experiments. However, it can be seen from the plotted node distributions that the coefficients of all three models fall below λ for scales outside the range $-3 \leq p \leq 3$. This discrepancy is due to the conservative nature of estimates of coefficient magnitudes $|c_j|_{\max}$, computed using equation (2.5).

The magnitudes of the errors in the network model's predictions of the light ball's position and velocity at a time 0.25s in the future are compared with those of a model of the object's dynamics consisting only of gravitational forces in figures 5-9 and 5-10. It can be seen that the predictions of the network model are more accurate than those of the non-adaptive model. Figures 5-11 and 5-12 give the corresponding prediction

errors of network and gravity models for the heavy ball experiment. The network model does not provide a significant improvement in prediction accuracy in this case since the only significant force experienced by the heavy ball was that due to gravity. The apparent convergence of errors in the non-adaptive model's predictions is a consequence of the transient response of the filter applied to numerically differentiated position measurements in the generation of velocity measurements.

Figures 5-13 and 5-14 compare the magnitudes of errors in predictions 0.25s in the future of the position and velocity of the paper aeroplane with those obtained assuming the object to have constant velocity over the prediction interval. Although variations in the attitude of the paper aeroplane over the course of the trajectory violate the assumptions used to derive the structure of the network model, the network-based identifier provides a clear improvement in prediction accuracy over the non-adaptive model. This can be explained as the result of high levels of noise in postion and velocity measurements, in addition to the inaccuracy of the constant velocity model.

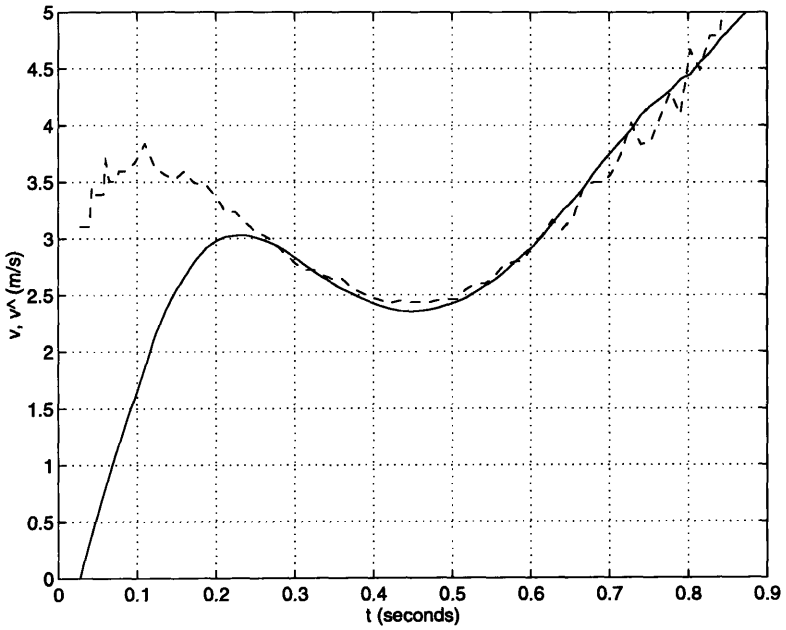

Figure 5-3: The convergence of the estimate $\hat{\mathbf{v}}(t)$ used to generate the error measure of the identifier of the light ball's dynamics. Solid line: $\|\hat{\mathbf{v}}(t)\|$. Broken line: the magnitude $\|\dot{\mathbf{x}}(t)\|$ of the measured velocity.

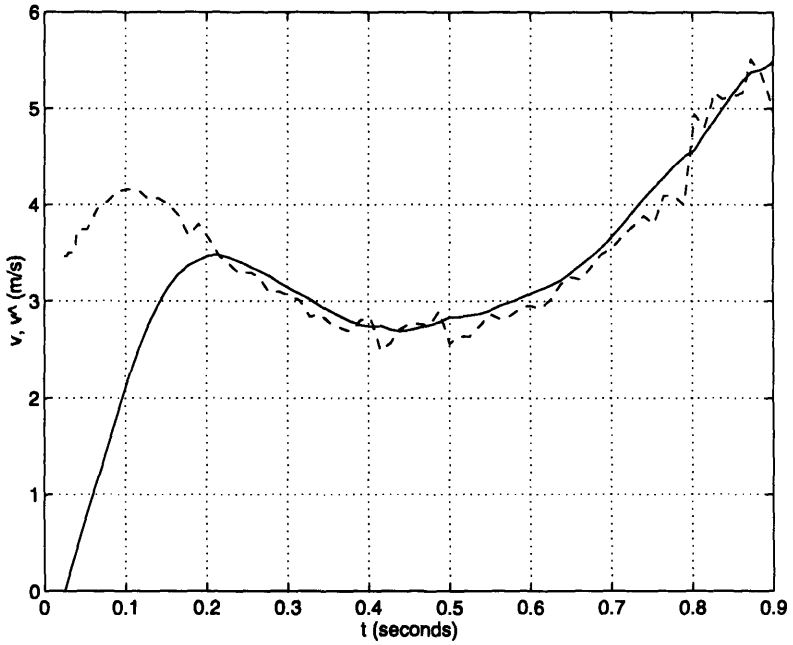

Figure 5-4: The convergence of the estimate $\hat{\mathbf{v}}(t)$ used to generate the error measure of the identifier of the heavy ball's dynamics. Solid line: $\|\hat{\mathbf{v}}(t)\|$. Broken line: the magnitude $\|\dot{\mathbf{x}}(t)\|$ of the measured velocity.

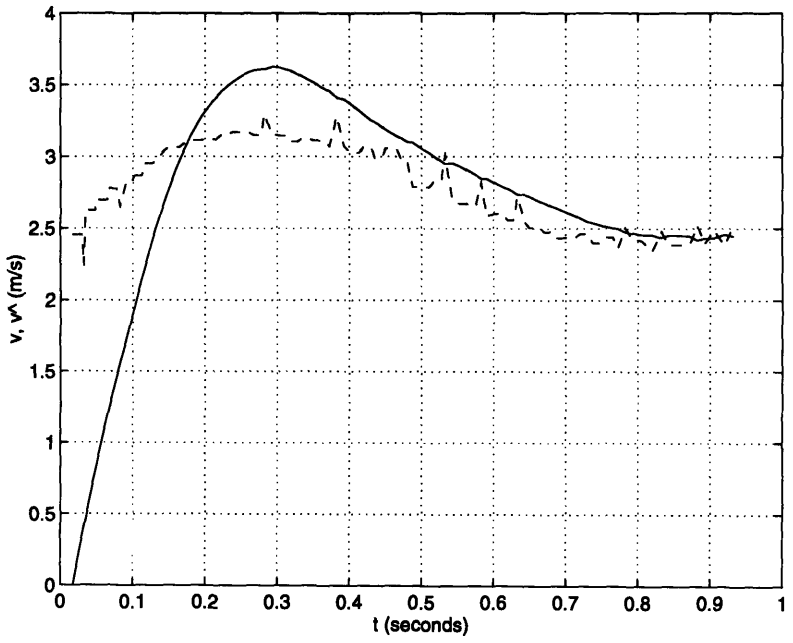

Figure 5-5: The convergence of the estimate $\hat{\mathbf{v}}(t)$ used to generate the error measure of the identifier of the paper aeroplane's dynamics. Solid line: $\|\hat{\mathbf{v}}(t)\|$. Broken line: the magnitude $\|\dot{\mathbf{x}}(t)\|$ of the measured velocity.

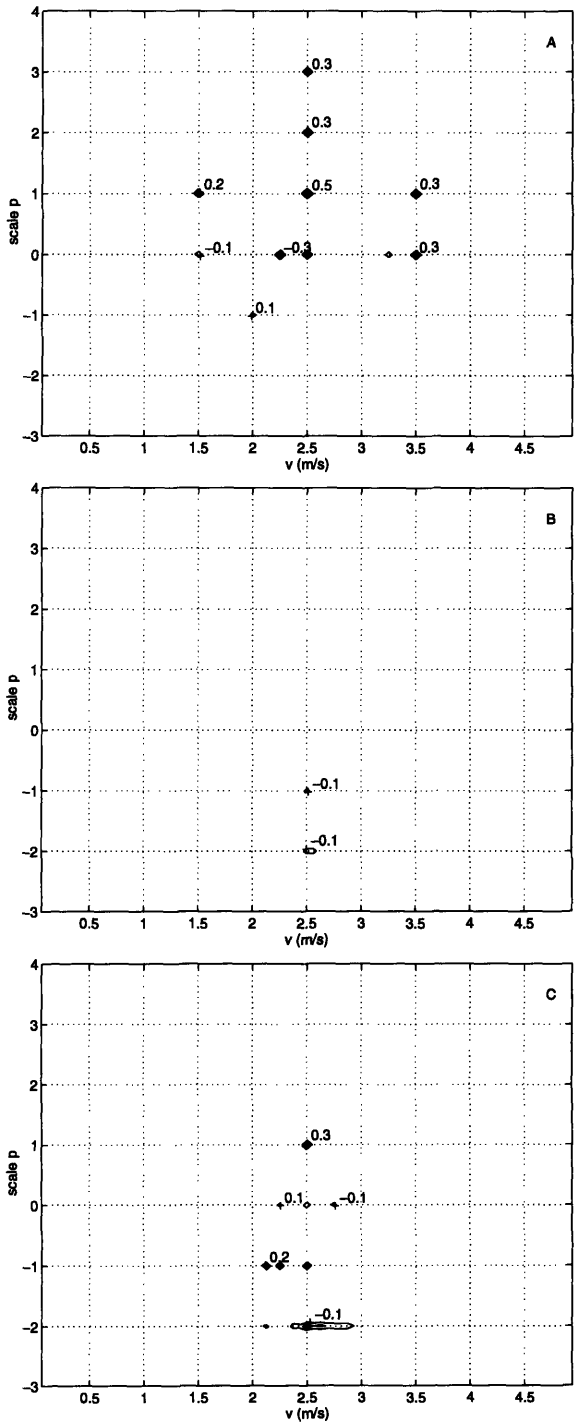

Figure 5-6: The coefficients and localization points in state-space (v) and logarithmic frequency (p) of the nodes of the networks generated in the experimental identification of a light ball's dynamics. A: network reconstructing $\hat{f}^t$ (28 nodes); B: network reconstructing $\hat{f}^{nc}$ (4 nodes); C: network reconstructing $\hat{f}^{ns}$ (13 nodes).

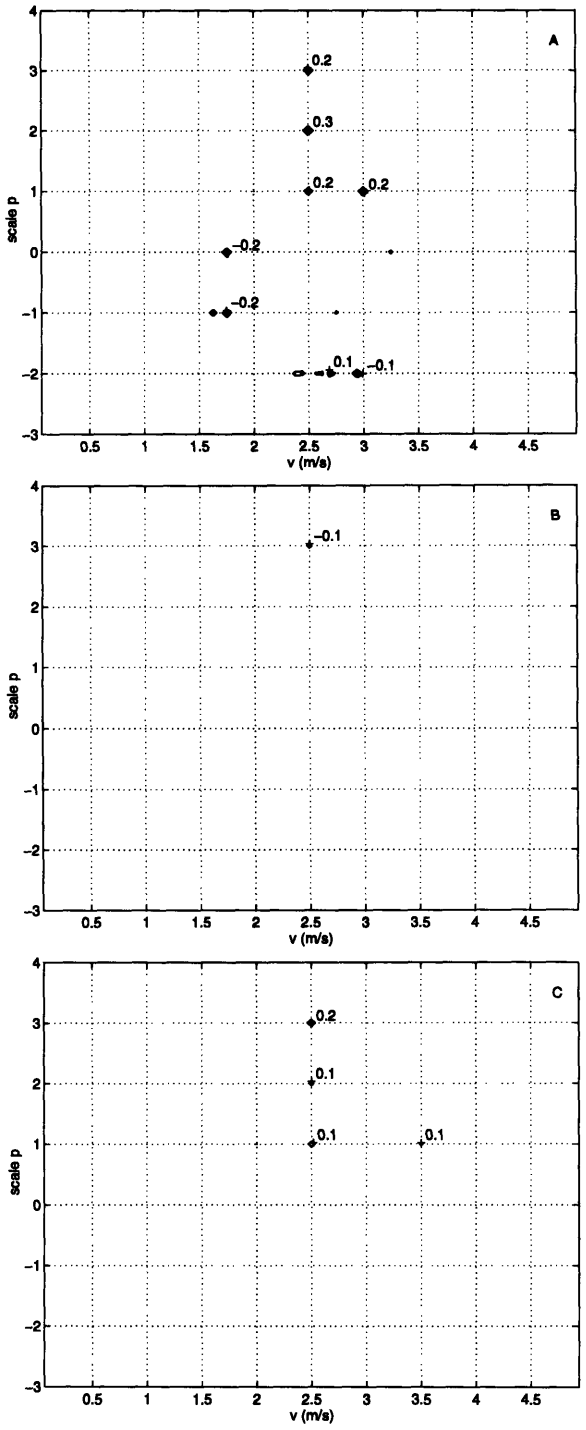

Figure 5-7: The coefficients and localization points in state-space (v) and logarithmic frequency (p) of the nodes of the networks generated in the experimental identification of a heavy ball's dynamics. A: network reconstructing $\hat{f}^t$ (21 nodes); B: network reconstructing $\hat{f}^{nc}$ (5 nodes); C: network reconstructing $\hat{f}^{ns}$ (6 nodes).

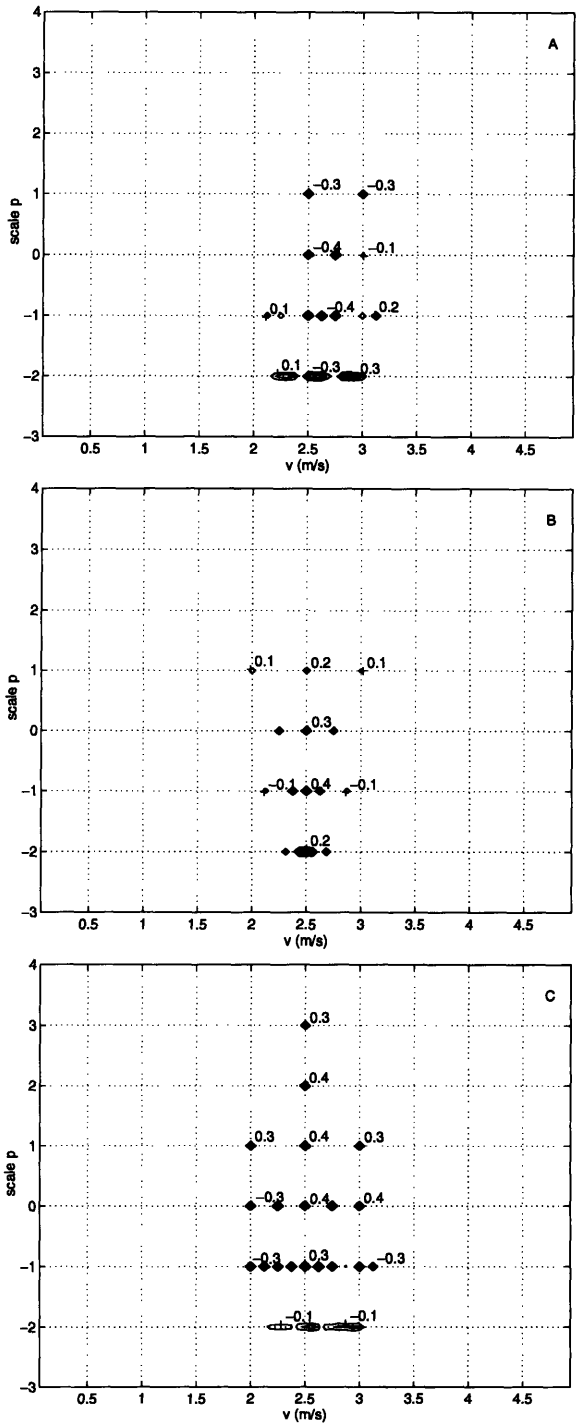

Figure 5-8: The coefficients and localization points in state-space (v) and logarithmic frequency (p) of the nodes of the networks generated in the experimental identification of a paper aeroplane's dynamics. A: network reconstructing $\hat{f}^t$ (21 nodes); B: network reconstructing $\hat{f}^{nc}$ (16 nodes); C: network reconstructing $\hat{f}^{ns}$ (34 nodes).

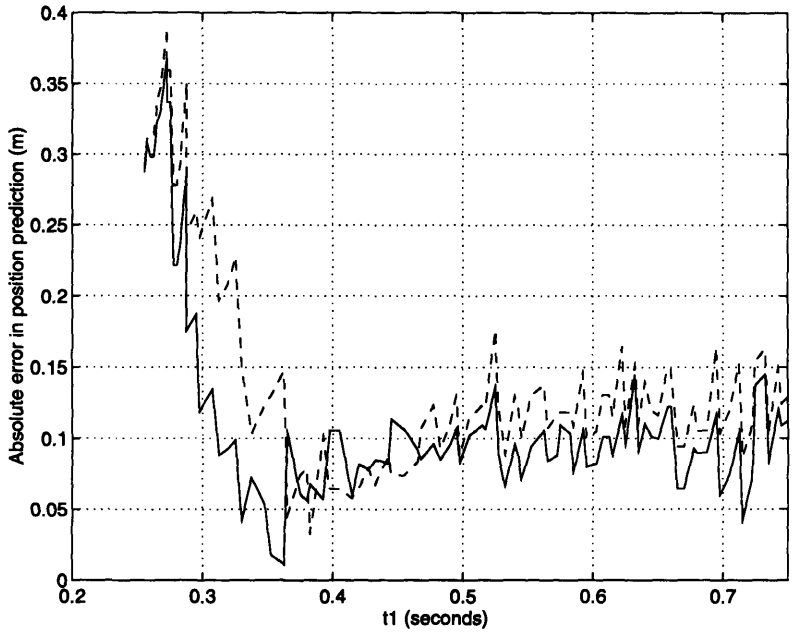

Figure 5-9: The absolute errors $\|\hat{\mathbf{x}}(t_1) - \mathbf{x}(t_1)\|$ in predictions $\hat{\mathbf{x}}(t_1)$, made at time t_0, of the position $\mathbf{x}(t_1)$ of a light ball at time t_1, where $t_1 - t_0 = 0.25$s. Solid line: network prediction errors. Broken line: errors in predictions based on gravitational acceleration.

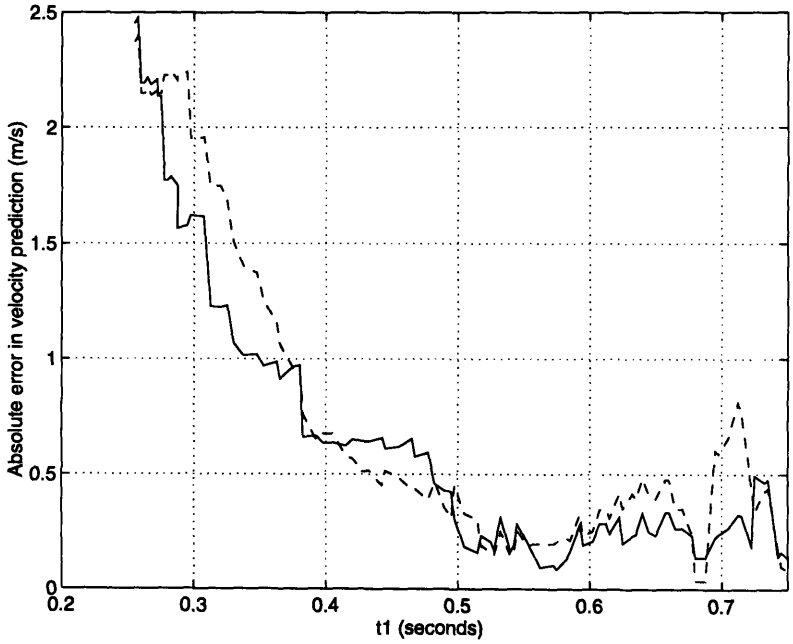

Figure 5-10: The absolute errors $\|\hat{\dot{\mathbf{x}}}(t_1) - \dot{\mathbf{x}}(t_1)\|$ in predictions $\hat{\dot{\mathbf{x}}}(t_1)$, made at time t_0, of the velocity $\dot{\mathbf{x}}(t_1)$ of a light ball where $t_1 - t_0 = 0.25$s. Solid line: network prediction errors. Broken line: errors in predictions based on gravitational acceleration.

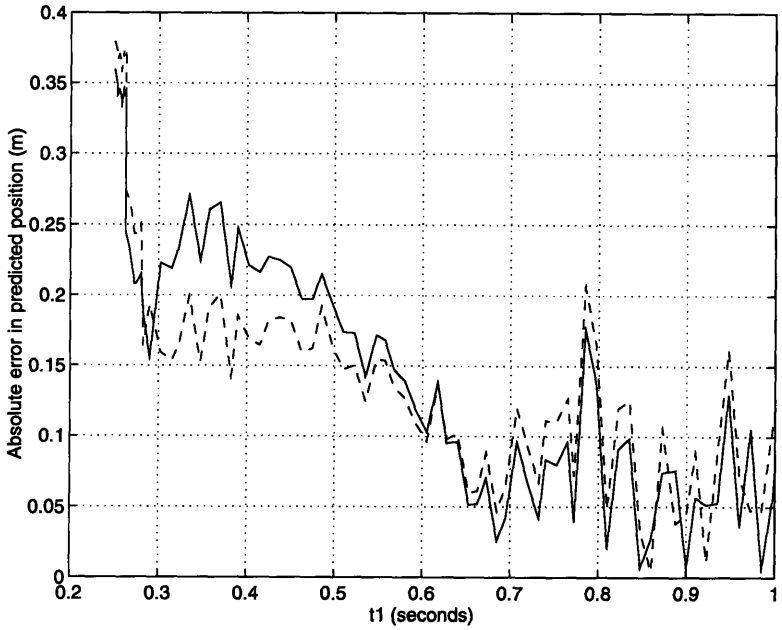

Figure 5-11: The absolute errors $\|\hat{\mathbf{x}}(t_1) - \mathbf{x}(t_1)\|$ in predictions $\hat{\mathbf{x}}(t_1)$, made at time t_0, of the position $\mathbf{x}(t_1)$ of a heavy ball at time t_1, where $t_1 - t_0 = 0.25$s. Solid line: network prediction errors. Broken line: errors in predictions based of gravitational acceleration.

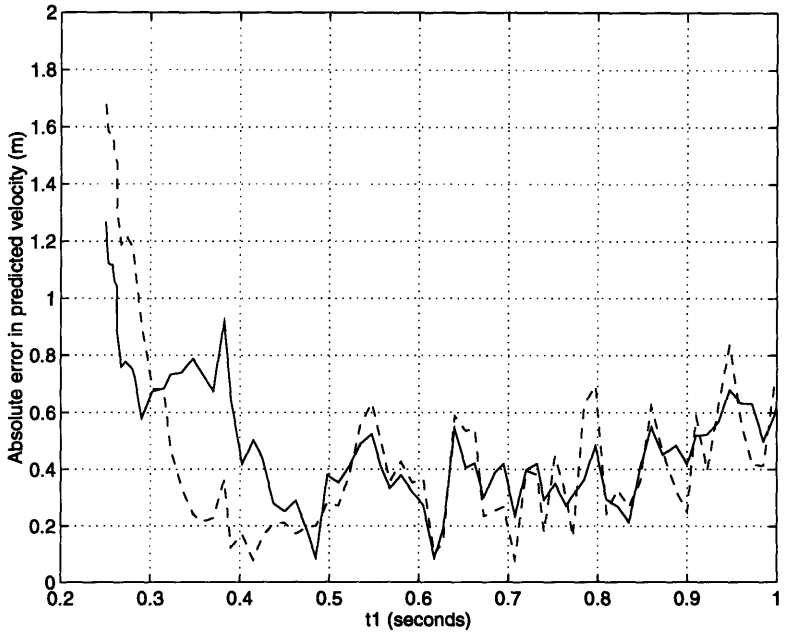

Figure 5-12: The absolute errors $\|\hat{\dot{\mathbf{x}}}(t_1) - \dot{\mathbf{x}}(t_1)\|$ in predictions $\hat{\dot{\mathbf{x}}}(t_1)$, made at time t_0 of the velocity $\dot{\mathbf{x}}(t_1)$ of a heavy ball where $t_1 - t_0 = 0.25$s. Solid line: network prediction errors. Broken line: errors in predictions based of gravitational acceleration.

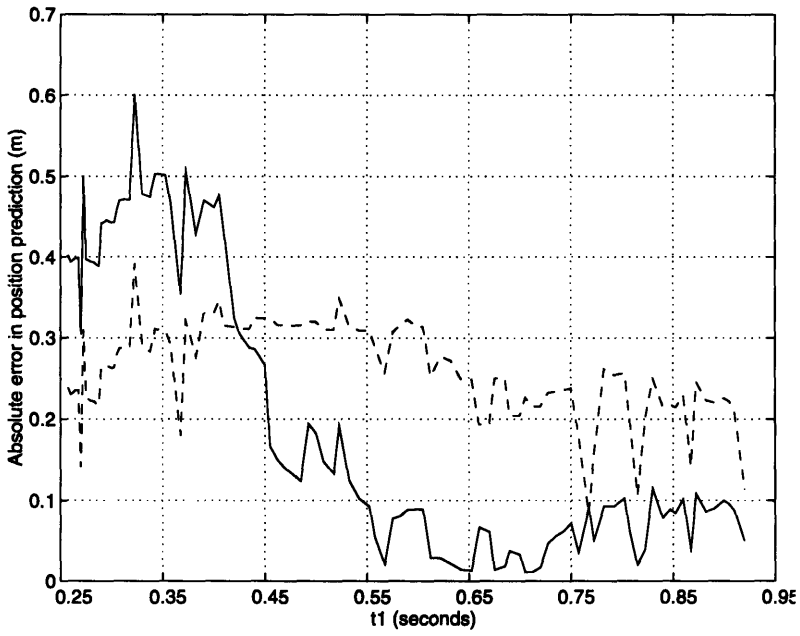

Figure 5-13: The absolute errors $\|\hat{\mathbf{x}}(t_1) - \mathbf{x}(t_1)\|$ in predictions $\hat{\mathbf{x}}(t_1)$, made at time t_0, of the position $\mathbf{x}(t_1)$ of a paper aeroplane at time t_1, where $t_1 - t_0 = 0.25$s. Solid line: network prediction errors. Broken line: errors in predictions made assuming constant velocity over the prediction interval.

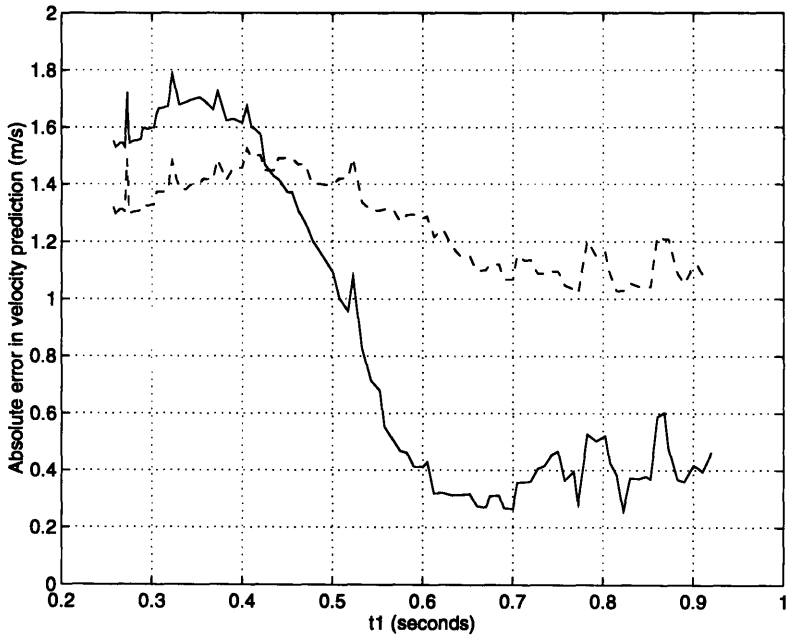

Figure 5-14: The absolute errors $\|\hat{\dot{\mathbf{x}}}(t_1) - \dot{\mathbf{x}}(t_1)\|$ in predictions $\hat{\dot{\mathbf{x}}}(t_1)$, made at time t_0 of the velocity $\dot{\mathbf{x}}(t_1)$ of a paper aeroplane where $t_1 - t_0 = 0.25$s. Solid line: network prediction errors. Broken line: errors in predictions made assuming constant velocity over the prediction interval.

Chapter 6

Concluding Remarks

A summary of the results presented within this thesis is as follows.

1. Algorithms for stably adapting on-line both the structure and output weights of a network consisting of a truncated expansion over a family of functions simultaneously localized in spatial and spatial frequency domains are developed.

2. On the basis of a bound on the temporal bandwidth of the controlled system, a criterion for determining the scalings of radial and ridge wavelet basis functions in the direction in state space of the derivative of the state vector is presented.

3. A criterion for selecting nodes for inclusion in the network expansion on the basis of thresholding rules applied to coefficient estimates is shown to minimize approximately the L^2-norm of the residual approximation error. For a class of non-orthonormal wavelet basis functions, it is shown that thresholding rules are optimal with respect to a regularization cost functional incorporating a smoothness measure that serves to quantify the degree to which the network's basis function set is matched to the wavelet transform of the approximated function.

4. An application of the network to the real-time identification of the dynamics of an unknown light bluff object thrown through air is described. The network model affords a higher degree of accuracy of predictions of the object's trajectory

over simple non-adaptive models of the object's dynamics if the aerodynamic forces experienced by the object are significant.

Bibliography

[1] A. Aronsajn. Reproducing kernel hilbert spaces. *Am. Math. Soc. Trans.*, 68:337–404, 1950.

[2] B. R. Bakshi and G. Stephanopoulos. Wave-net: A multiresolution, hierarchical neural network with localized learning. *AIChE J.*, 39(1), 1993.

[3] A. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inform. Theory*, 39:930–945, May 1993.

[4] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel And Distributed Computation.* Prentice Hall, Englewood Cliffs, NJ, 1989.

[5] K. C. Chou, A. S. Willsky, and A. Benveniste. Multiscale recursive estimation, data fusion, and regularization. Technical Report LIDS-P-2085, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 1991.

[6] G. Cybenco. Approximations by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 309(2), 1989.

[7] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inform. Theory*, 36:961–1005, September 1990.

[8] I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF regional conference series in applied mathematics, 1992.

[9] I. Daubechies, A. Grossmann, and Y. Meyer. Painless non-orthogonal expansions. *J. Math. Phys.*, 27:1271–1283, 1986.

[10] B. Delyon, P-Y. Glorennec, A. Juditsky, Q. Zhang, and A. Benveniste. Wavelets in identification. Technical report, IRISA, April 1994.

[11] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. Technical report, Department of Statistics, Stanford University, 1992.

[12] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. Technical report, Department of Statistics, Stanford University, 1992.

[13] N. Dyn. Interpolation of scattered data by radial functions. In C. K. Chui, L. L. Schumaker, and D. J. Ward, editors, *Topics in Multivariate Approximation.* Academic Press, New York, 1987.

[14] N. Dyn. Interpolation and approximation by radial and related functions. In C. K. Chui and D. J. Ward, editors, *Approximation Theory VI.* Academic Press, New York, 1991.

[15] F. Girosi. Some extensions of radial basis functions and their applications in artificial intelligence. *Computers Math. Applic.*, 24(12):61–80, 1992.

[16] F. Girosi, M. Jones, and T. Poggio. Priors, stabilizers, and basis functions: from regularization to radial, tensor and additive splines. Technical report, AI Memo No. 1430, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, June 1993.

[17] F. Girosi and T. Poggio. Networks for approximation and learning. *Proc. IEEE*, 78(9):1481–1497, September 1990.

[18] H. K. Khalil. *Nonlinear Systems.* Macmillan, 1992.

[19] M. R. Luettgen, W. C. Karl, and A. S. Willsky. Optical flow computation via multiscale regularization. Technical Report LIDS-P-2115, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, June 1992.

[20] W. R. Madych and Nelson S. A. Multivariate interpolation and conditionally positive definite functions ii. *Math. Comp.*, 54(189):211–230, January 1992.

[21] S. Mallat. *Wavelet Signal Processing.* Academic Press, New York, 1995.

[22] S. Mallat and W. L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. Inform. Theory*, 38:617–643, 1992.

[23] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. Technical Report 619, New York University, Computer Science Department, August 1993.

[24] R. Martin and D. Cochran. Generalized wavelet transforms and the cortex transform. *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, November 1994.

[25] Y. Meyer. *Ondelettes et Operateurs.* Hermann, 1990.

[26] Y. Meyer and S. Rocques, editors. *Progress In Wavelet Analysis And Applications: Proceedings Of The International Conference "Wavelets And Applications", Toulouse, France*, June 1992.

[27] C. A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constr. Approx.*, 2:11–22, 1986.

[28] R. Murenzi. The wavelet transforms associated to the n-dimensional Euclidean group with dilations: Signals in more than one dimension. In J. M. Combes, A. Grossmann, and Ph. Tchamitchian, editors, *Wavelets*, pages 239–246. Springer-Verlag, Berlin, 1989.

[29] R. Murenzi. *Ondelettes multidimensionelles et application à l'analyse d'images.* PhD thesis, Université Catholic de Louvain, Belgium, 1990.

[30] P. Niyogi and F. Girosi. On the relationship between generalization error, hypothesis complexity and sample complexity for radial basis functions. Technical

report, AI Memo No. 1467, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, February 1994.

[31] M. Polycarpou and P. Ioannou. Identification and control of nonlinear systems using neural network models: Design and stability analysis. Technical Report 91-09-01, USC Dept. EE-Systems, September 1991.

[32] C. S. Rees, S. M. Shah, and Stanojević Ć. V. *Theory And Applicayions Of Fourier Analysis.* Marcel Dekker, New York, 1981.

[33] R. Sanner. *Stable adaptive control and recursive identification of nonlinear systems using radial gaussian networks.* PhD thesis, Massachusetts Institute of Technology, 1993.

[34] R. Sanner and J-J. Slotine. Gaussian networks for direct adaptive control. *IEEE Trans. Neural Networks*, November 1992.

[35] R. Sanner and J-J. Slotine. Stable adaptive control of robot manipulators using 'neural' networks. *Neural Computation*, 1994.

[36] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Process.*, 41(12):3445–3462, December 1993.

[37] J-J. Slotine and W. Li. *Applied Nonlinear Control.* Prentice Hall, Englewood Cliffs, NJ, 1991.

[38] H. H. Szu, B. Telfer, and S. Kadambe. Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31(9):1907–1916, September 1992.

[39] G. Wahba. *Spline Models For Observational Data.* SIAM Series In Applied Mathematics, vol. 59, SIAM Philadelphia, 1990.

[40] Q. Zhang. Regressor selection and wavelet network construction. *Proc. 32nd IEEE Conference on Decision and Control*, pages 3688–3693, 1993.